

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:59:31 ; Search time 15.6054 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-2

Perfect score: 3810
Sequence: 1 TDRVSGNLRARVLYDFVN.....KARAEEKPAPDRAGDDAAR 740

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*\n2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*\n3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*\n4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*\n5: /cgn2_6/ptodata/1/1aa/PTCUS-COMB.pep:*\n6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3810	100.0	741	3	US-09-001-984C-106
2	3810	100.0	741	4	US-09-396-347F-106
3	2570	67.5	731	4	US-09-252-991A-22524
4	2289	60.1	739	2	US-08-836-943-2
5	2225	58.4	729	4	US-09-328-352-7603
6	139	3.6	3892	4	US-09-328-352-5503
7	135	3.5	3816	3	US-09-428-517-3
8	128	3.4	878	1	US-08-237-919-2
9	128	3.4	878	4	US-08-732-429-2
10	128	3.4	878	4	US-09-798-267-2
11	128	3.4	878	4	US-09-798-267-3
12	128	3.4	878	5	PCT-US95-05518-2
13	128	3.4	11877	3	US-09-105-537-6
14	127	3.3	1850	4	US-09-252-991A-21798
15	126.5	3.3	3739	3	US-09-320-878-2
16	126.5	3.3	3739	3	US-09-105-537-33
17	126.5	3.3	3739	4	US-09-141-908-3
18	126.5	3.3	3739	4	US-09-657-440-2
19	125.5	3.3	2039	4	US-09-077-098A-7
20	124	3.3	1346	3	US-09-105-537-37
21	124	3.3	15281	2	US-08-471-119A-2
22	123	3.2	1346	3	US-09-320-878-4
23	123	3.2	1346	4	US-09-141-908-5
24	123	3.2	1346	4	US-09-657-440-4
25	123	3.2	1529	4	US-09-657-440-5
26	120.5	3.2	675	4	US-09-134-001C-3945
27	119.5	3.1	2032	4	US-09-071-035-458

28	119.5	3.1	2032	4	US-09-071-035-462	Sequence 452, Appl
29	119.5	3.1	2032	4	US-09-071-035-466	Sequence 456, Appl
30	119.5	3.1	7257	3	US-08-335-409-5	Sequence 5, Appl
31	119.5	3.1	7257	4	US-08-568-102-5	Sequence 5, Appl
32	119.5	3.1	7257	4	US-09-567-969-5	Sequence 5, Appl
33	119.5	3.1	7257	4	US-09-568-480-5	Sequence 5, Appl
34	119.5	3.1	7257	4	US-09-568-486-5	Sequence 5, Appl
35	119.5	3.1	7257	4	US-09-568-472-5	Sequence 5, Appl
36	119.5	3.1	7257	4	US-09-567-899-5	Sequence 5, Appl
37	118.5	3.1	915	4	US-09-206-942-35	Sequence 35, Appl
38	118.5	3.1	1222	4	US-09-206-942-37	Sequence 37, Appl
39	118.5	3.1	1228	4	US-09-206-942-34	Sequence 34, Appl
40	118	3.1	1385	4	US-09-252-991A-21919	Sequence 21919, A
41	116.5	3.1	1289	2	US-08-853-659A-51	Sequence 51, Appl
42	116.5	3.1	4928	3	US-09-036-987A-5	Sequence 5, Appl
43	116.5	3.1	4928	3	US-09-370-700-5	Sequence 5, Appl
44	116.5	3.1	4928	4	US-09-603-207-5	Sequence 5, Appl
45	115	3.0	2042	4	US-09-077-098A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-001-984C-106
; Sequence 106, Application US/09001984C
; Patent No. 6245331
; GENERAL INFORMATION:
; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Belisle, John T
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: NYU-011
; CURRENT APPLICATION NUMBER: US/09/001.984C
; CURRENT FILING DATE: 1997-12-31
; PRIOR APPLICATION NUMBER: 60/034,003
; PRIOR FILING DATE: 1996-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-001-984C-106

Query Match		100.0%;	Score 3810;	DB 3;	Length 741;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 740;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TDRVSGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ	60		
	2	TDRVSGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ	61		
Db	61	AQIDKWHRRVIBPDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV	120		
	62	AQIDKWHRRVIBPDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV	121		
Qy	121	LNARFALNANARWSLYDALYCTDVIPTDGAKEGTYNKGDKVIARXKFLDSDVP	180		
	122	LNARFALNANARWSLYDALYCTDVIPTDGAKEGTYNKGDKVIARXKFLDSDVP	181		
Qy	181	LSSGSFGDATGFTVQDGLVVALPKSTGLANPQAGYTGAAESPTSVLLINRGLHIEI	240		
	182	LSSGSFGDATGFTVQDGLVVALPKSTGLANPQAGYTGAAESPTSVLLINRGLHIEI	241		
Qy	241	LIPESQVGTTRAGVKDVILESAITTIMDFESVAAVDAADKVLGVRNWLGNKGLAA	300		
	242	LIPESQVGTTRAGVKDVILESAITTIMDFESVAAVDAADKVLGVRNWLGNKGLAA	301		
Qy	301	AVDKDGFARVLNRDRNTAPGGQFTLPGRSLMFRVNVGHMTNDALVDTDGSEVFEG	360		
	302	AVDKDGFARVLNRDRNTAPGGQFTLPGRSLMFRVNVGHMTNDALVDTDGSEVFEG	361		

361 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPFAEVAFTCELFSEVEDVLG 420
Db
362 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPFAEVAFTCELFSEVEDVLG 421
Qy
421 LPQNTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMVRKG 480
Db
422 LPQNTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMVRKG 481
Qy
481 TMSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADWVETKIAQPRAGASTAW 540
Db
482 TMSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADWVETKIAQPRAGASTAW 541
Qy
541 VPSPATAATHALHYHVDVAAVQOGLAGKRATIEQLLTIPLAKELAWAPDEIREVEDNN 600
Db
542 VPSPATAATHALHYHVDVAAVQOGLAGKRATIEQLLTIPLAKELAWAPDEIREVEDNN 601
Qy
601 COSILGYVVRWVDQGVCSKVPDIHDVALMEDRATLRISOLLANWLHGVITSADVRAS 660
Db
602 COSILGYVVRWVDQGVCSKVPDIHDVALMEDRATLRISOLLANWLHGVITSADVRAS 661
Qy
661 LERMAPLVDRONAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGYTEPILHRRRREF 720
Db
662 LERMAPLVDRONAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGYTEPILHRRRREF 721
Qy
721 KARAAEKPAAPSDRAGDDAAR 740
Db
722 KARAAEKPAAPSDRAGDDAAR 741
RESULT 2
US-09-396-347F-106
; Sequence 106, Application US/09396347F
; Patent No. 6506384
; GENERAL INFORMATION:
; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Bellis, John T
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: 32004-169276
; CURRENT APPLICATION NUMBER: US/09/396,347F
; CURRENT FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: 09/001,984
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-396-347F-106
Query Match 100.0%; Score 3810; DB 4; Length 741;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNOALLNARDELQ 60
Db 2 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNOALLNARDELQ 61
Qy 61 AOIDKWHRRRVEIPIDMDAYEQFTEIGYLLPEPDDFTITTSVGDAEITTTAGPQVVPV 120
Db 62 AOIDKWHRRRVEIPIDMDAYEQFTEIGYLLPEPDDFTITTSVGDAEITTTAGPQVVPV 121
Qy 121 LNARFALNAANARWGSGLYDALYGTVDVPEIDGAEKGYTNKVRGDKVIAYARKFLDSDVP 180
Db 122 LNARFALNAANARWGSGLYDALYGTVDVPEIDGAEKGYTNKVRGDKVIAYARKFLDSDVP 181
Qy 181 LSSGSFGDGTFTVDQGVVALPKSTGLANPQFAGYTCABSPSVLLINHLHIEI 240
Db 182 LSSGSFGDGTFTVDQGVVALPKSTGLANPQFAGYTCABSPSVLLINHLHIEI 241
Qy 241 LIDPESQVGTTRDAGVKDVILESATITIMDFEDSVAAVDAADKVLGYRNWMLGNKGLDAA 300

242 LIDPESQVGTTRDAGVKDVILESATITIMDFEDSVAAVDAADKVLGYRNWMLGNKGLDAA 301
Qy
301 AVDXDGTAFLEVLNRDRNYTAPGGGQFTLPGRSIMFVRNVGHLMTNDAIYDITDGSSEVFE 360
Db
302 AVDXDGTAFLEVLNRDRNYTAPGGGQFTLPGRSIMFVRNVGHLMTNDAIYDITDGSSEVFE 361
Qy
361 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPFAEVAFTCELFSEVEDVLG 420
Db
362 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPFAEVAFTCELFSEVEDVLG 421
Qy
421 LPQNTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMVRKG 480
Db
422 LPQNTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMVRKG 481
Qy
481 TMSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADWVETKIAQPRAGASTAW 540
Db
482 TMSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADWVETKIAQPRAGASTAW 541
Qy
541 VPSPATAATHALHYHVDVAAVQOGLAGKRATIEQLLTIPLAKELAWAPDEIREVEDNN 600
Db
542 VPSPATAATHALHYHVDVAAVQOGLAGKRATIEQLLTIPLAKELAWAPDEIREVEDNN 601
Qy
601 COSILGYVVRWVDQGVCSKVPDIHDVALMEDRATLRISOLLANWLHGVITSADVRAS 660
Db
602 COSILGYVVRWVDQGVCSKVPDIHDVALMEDRATLRISOLLANWLHGVITSADVRAS 661
Qy
661 LERMAPLVDRONAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGYTEPILHRRRREF 720
Db
662 LERMAPLVDRONAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGYTEPILHRRRREF 721
Qy
721 KARAAEKPAAPSDRAGDDAAR 740
Db
722 KARAAEKPAAPSDRAGDDAAR 741
RESULT 3
US-09-252-991A-22524
; Sequence 22524, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22524
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22524
Query Match 67.5%; Score 2570; DB 4; Length 731;
Best Local Similarity 69.1%; Pred. No. 2.1e-240;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;
Qy 1 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNOALLNARDELQ 60
Db 8 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNOALLNARDELQ 67
Qy 61 AOIDKWHRRRVEIPIDMDAYEQFTEIGYLLPEPDDFTITTSVGDAEITTTAGPQVVPV 120
Db 68 AOIDKWHRRRVEIPIDMDAYEQFTEIGYLLPEPDDFTITTSVGDAEITTTAGPQVVPV 127
Qy 121 LNARFALNAANARWGSGLYDALYGTVDVPEIDGAEKGYTNKVRGDKVIAYARKFLDSDVP 180
Db 128 MNARFALNAANARWGSGLYDALYGTVDVPEIDGAEKGYTNKVRGDKVIAYARKFLDSDVP 187

QY 181 LSSGSGDATGFTVQDQOLVWALPKS-TGLANPGQFAGYTGAAESPTSVLLINHLHIE 239
Db 188 LESGSHVDATSYVKNGALVKNKNGSETGLKNAGQFLAQDGAAPQAVLKNHGLHFE 247
QY 240 ILIDPESQVGTDRAGVKDVLKESAITTIMDPEDSVAAVDAADKVLGYRNWGLNKGDLA 299
Db 248 IQIDPSSPVQGTDAAGVKDVLKESAITTIMDCEDSVAAVDAADKVVYRNWGLNKGDLA 307
QY 300 AAVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSLMFVRNVLHMTNDAIVTDTGSEVFE 359
Db 308 EYVSKGSGTFTTMRNPDVYTRADGSELTHGRSLLFVRNVLHMTNDAIILDKDGNVPE 367
QY 360 GIMDALFTGLIAHGLKASDVNGPLI--NSRTGSIIVIVKPKMHGPAEVAFTCELSRYED 417
Db 368 GIDQGLFTSLIAH-----DLNGNSRKNSRTGSIIVIVKPKMHGPEEAAFTNELFRVED 422
QY 418 VLGLPONTWKIGIMDEERTTNVNLKACIKAAADRVVFINTEGLDRGTGSIHTSMAGPMV 477
Db 423 VLGLPNTLKVGTIMDEERTTNVNLKACIKAAADRVVFINTEGLDRGTGSIHTSMAGAVV 482
QY 478 RGTMTKSQPKILAYEDHNVDAAGLAAFGSRAQVKGKMTMTLMADMVETKIAQPRAGAS 537
Db 483 RKGAMKSEKIGAYENNNVDVGLATGLQKQAQIGKGMWAMPDLMAANLEQKIGHPLAGAN 542
QY 538 TAWVPSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLTIPLAKELAWAPDEIRBEV 597
Db 543 TAWVPSPTAATLHALHYHVDVFAQAELAKTRPASVDLTIPLAPNTNWTAEIKNEV 602
QY 598 DNNQCSILGYVVRWVDQVGCSCVPDIHDVVALMEDRATLRISSOLLANLWLRHGVITSADV 657
Db 603 DNNQCSILGYVVRWVDQVGCSCVPDIHDVVALMEDRATLRISSOLLANLWLRHGVISQEV 662
QY 658 RASLERMAPLVRQNAQDVAYRPMAPNFDSDIAFLAAQELIISGAQQQNGYTEPILHRRR 717
Db 663 VESLKEMAVVRQNASDPSYRPMAPNFDNDVAFQAALELVVEGTQNGYTEPILHRRR 722
QY 718 REFKA 723
Db 723 REFKA 728

RESULT 4

US-08-836-943-2
; Sequence 2, Application US/08836943
; Patent No. 5965391
; GENERAL INFORMATION:
; APPLICANT: Reinscheid, Dieter
; APPLICANT: Bickmanns, Bernhard
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: DNA WHICH REGULATES GENE EXPRESSION IN
; TITLE OF INVENTION: CORYNEFORM BACTERIA
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSES: The Firm of Karl F. Ross, PC
; STREET: 5676 Riverdale Ave.
; CITY: Bronx
; STATE: New York
; COUNTRY: USA
; ZIP: 10471
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836.943
; FILING DATE: 08-MAY-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Jonathan
; REGISTRATION NUMBER: 26,963
; REFERENCE/DOCKET NUMBER: 20357
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (718) 884-6600
; TELEFAX: 718/601-1099
; TELEX: 620428
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 739 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-836-943-2

Query Match 60.1%; Score 2289; DB 2; Length 739;

Best Local Similarity 61.3%; Pred. No. 4.1e-213;
Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;
QY 1 TDRVSVGNLRIRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTTFQNALNARDELQ 60
Db 20 TERVDAGMQVAKVLYDFVTEAVLPRVGVDAEKFWSGFAAARDLTPRNEALLARDELQ 79
QY 61 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPBPDDTITTSQVDAEITTAGPOLVVPV 120
Db 80 MLIDDYHRNN-SGTIDQAYEDFLKEIGYLVVEEPEAAEIRTONVDTEISSTAGPOLVVPV 138
QY 121 LNARFALNANARWGLYDALYGTDPETDGAEGKPTYNKVRGDKVIAYARFDDSV 180
Db 139 LNARFALNANARWGLYDALYGTNAIPETDGAEGKPTYNKVRGDKVIAYARFDDSV 198
QY 181 LSSGSGDATGFTVQDQOLVWALPKS-TGLANPGQFAGYTGAAESPTSVLLINHLHIE 240
Db 199 LDGASHADVEKYNITDCKLAHIGDSVYRLKNRESYRGFTGNFLDPEAILLETNGLHIE 258
QY 241 LIDPESQVGTDRAGVKDVLKESAITTIMDPEDSVAAVDAADKVLGYRNWGLNKGDLA 300
Db 259 QIDVHPHIGKADTKGLKDIIVLESALITTIMDPEDSVAAVDAADKVLGYRNWGLNKGDLA 318
QY 301 AVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSLMFVRNVLHMTNDAIVTDTGSEVFE 360
Db 319 EMSKNGRIFTRELKORVYIGRNGTELVLHGRSILFVRNVLHMTNDAIVTDTGSEVFE 377
QY 361 IMDALFTGLIAHGLKASDVNGPLI--NSRTGSIIVIVKPKMHGPAEVAFTCELSRYED 420
Db 378 IMDAVLTVCALPGIAFQN---KMRNSRKSIIIVKPKMHGPEEAAFTNELFRVED 434
QY 421 LPQNTMKIGIMDEERTTNVNLKACIKAAADRVVFINTEGLDRGTGSIHTSMAGPMV 480
Db 435 LPRHTLVKGVMDBERTSVNLDAIMEVADRLAFINTGFLDRGTGSIHTSMAGPMV 494
QY 481 TMKSQPMILAYEDHNVDAAGLAAFGSRAQVKGKMTMTLMADMVETKIAQPRAGASTAW 540
Db 495 DMQTAPWKQAYENNVDAAGLAAFGSRAQVKGKMTMTLMADMVETKIAQPRAGASTAW 554
QY 541 VPSPTAATLHALHYHVDVAAVQOGL-AGKERATIEQLTIPLAKELAWAPDEIRBEVDN 599
Db 555 VPSPTAATLHALHYHVDVAAVQOGL-AGKERATIEQLTIPLAKELAWAPDEIRBEVDN 614
QY 600 NCQCSILGYVVRWVDQVGCSCVPDIHDVVALMEDRATLRISSOLLANLWLRHGVITSADV 659
Db 615 NCQCSILGYVVRWVDQVGCSCVPDIHDVVALMEDRATLRISSOLLANLWLRHGVITSADV 674
QY 660 SLERMAPLVRQNAQDVAYRPMAPNFDSDIAFLAAQELIISGAQQQNGYTEPILHRRR 719
Db 675 SLERMAPLVRQNAQDVAYRPMAPNFDSDIAFLAAQELIISGAQQQNGYTEPILHRRR 734
QY 720 FKAR 723
Db 735 FKAR 738

RESULT 5

US-09-328-352-7603
; Sequence 7603, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:

US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; FILE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3816
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-428-517-3
Query Match 3.5%; Score 135; DB 3; Length 3816;
Best Local Similarity 21.3%; Pred. No. 0.01;
Matches 190; Conservative 112; Mismatches 297; Indels 294; Gaps 49;
QY 25 PG-----TDTPD-SFWAGVDKVVADLTTPQNALNARDLQAIQDKHRRVIEPID---- 76
DB 2968 PGRFVLADIGDGBASDMLPRAVA-----SAASEVIRAGAVVYPRLARATDEGL 3017
QY 77 -----MDAYRQLTIG-----YLLPEPDPTITTSVGDABETTTTAGPOLVVPVLNARF 125
DB 3018 VVADEAAGPWRLDVTAGILANLALVPCD-----ASRPLGPEV-----RI 3059
QY 126 ALNANARNGSLYDALYGTVDIPETD--GAEGKPTYNKV-----RGKVIAYARKEL 175
DB 3060 AVRAAGV--NFRDVLALGMYPDEGLMGAEEAGVTEVGGVTTILAPGDRVWG----- 3110
QY 176 DDSVPLSSGSGFSGATGFTVQDGLVVALPKDSTGLANPGCFAGYTGAAESPTSVLLINH 235
DB 3111 -----LVTFGFGV---AVTHRRMLVRMP-----RGWSFAEASVPVAFLLAYYA 3152
QY 236 LH-----IEILIDPESQVGTDRAGVKDVI-----LE 262
DB 3153 LHDLAGLRGESVLVHSAAGVGMAAVALARHWDAAEVFGTASGKWDVLAQAQGLDEHIG 3212
QY 263 SAITIMDFEDSVAA-----VDAADKVL--GYRNWGLNKGDL--A 299
DB 3213 SSRTT--EPEQRPATSGRGHIDVNLNLSGDFVDASARLLREGGR-FVEMGKTDIRDL 3269
QY 300 AAVDKDGTAFRLVRNDRNVTAPGG-----GQFTLP-----GR 332
DB 3270 GVVGADGVPIRVAVD---LAEGAERIGQMLDEIMALFDAGVLELPLRANVPVRAHE 3326
QY 333 SLMFV---RNUGH-LMNTDAIVDTGSEVPFGIMDALFTGLIAHGLKASDVNGPLNRS 388
DB 3327 ALRFVSGARHGKRWLTVPAALDAEGTVLTGA--CTLGALVARHLVTEHDEVRLLLYSR 3384
QY 389 TGSIIYVKPMHGP-----AEVAF--TCELFSR-----VEDVLGLPQNTM-----KIGIM 431
DB 3385 SG-----VAPDLAELGALGAETVVAACDVANKALKALLEDI--PPEHPTGIVHTAGVL 3438
QY 432 DE-----ERRTIVNLKACIAAADRVVFFINTGFLDRTGDEIHT---SMEAGPMVRKG 480
DB 3439 DCGWSGLPVRDVT-LXPKVDAALTLESVIGELDLDPALFVITSSAASMLGGPQGSY 3497
QY 481 TMKSQPIILAYEDHNVDAAGAFSGRAQVGKMW-----TMTLMADWVETKIAOPRAGA 536

DB 3498 AAANQ-FLDTLARHRRRGLTS-----VSLGWGLWHEASGLTGGIADIDRDMSS--RAGI 3549
QY 537 STAWVPSTAAHL-----ALHYHOVDVAVAOQGLAGKRRATIEOLL----- 578
DB 3550 A-----PMPTDEALHLFDRATELGDPVLLPMLNEAALEDRAAD---GTLPLLSGLVVR 3602
QY 579 -----TIP--LAKELAWAPD---BIRBEVDNCCOSILGYV-----VWV 612
DB 3603 HRPSARAGTATAATGPEAFARLAAAPDPRRALDLVRGHVALVLCHSGPEAIDAEQA 3662
QY 613 DQGVCSKVPIHDVVALMEDRATIRISSQLANWLHGVITSADVRSLEMAPLVDRQN 672
DB 3663 FRDIGFDSLTAVELNRNLNAETGLRLPQTLVFDYPN-----PSALADHLELLAP----- 3712
QY 673 AGDVAYREMAFNFDSDIAFLAAQELILSGAQOQNGYTEPILHRRRRERFKARAA 725
DB 3713 -----ATQTPAAL---JAELEERVEQLLSAAASPGPASAVDEETRTLIATRIA 3758
RESULT 8
US-08-237-919-2
; Sequence 2, Application US/08237919
; Patent No. 5610281
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B
; APPLICANT: Cepek, Karyn L
; TITLE OF INVENTION: Methods and Compositions for
; TITLE OF INVENTION: Modulating Heterotypic E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/237,919
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-237-919-2
Query Match 3.4%; Score 128; DB 1; Length 878;
Best Local Similarity 22.0%; Pred. No. 0.004;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;
QY 91 LPEDDDFTI-----TTSVDABETTTTAGPOLVVPVNLNARFALNANARWGLYDAL 141
DB 307 LPDXNMTINRNTGVISVITGLDRESFPTV--TLVV-----QAADIQSGGL--ST 353
QY 142 YGTDVIVPETDGAEGKPTYNKV--RGD-----KVIAVARKFLDSDVPLSSGSGFCDATGPT 193
DB 354 TATAVITVTDNDNPPFNPTTYKQVPENAEANVVITLKVTDADAPNTPAWEAVYILN 413
QY 194 VQDGLVVAL-PDKSTGLAN-----PGQFAGVTGAASPTSV 229

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-732-429-2

Query Match
Best Local Similarity 3.4%; Score 128; DB 4; Length 878;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY 91 LPDPDFTI-----TTSQVDAEITTTAGPOLVVPVFNARFALNAARWGSLSYDAL 141
DB 307 LPDKMFTINENTGVISVVTTLGDRSEPTV--TLVV-----QAADIQSGSL--ST 353
QY 142 YGTVIPIETDGAEGKPTNKV--RGD-----KVIAYARKFLDDSVPLSSGSGFDATGFT 193
DB 354 TATAVITVTNDNPNPIFNFTYKQVPEANEANVVITLKVTDADAPNPAWEAVYITLN 413
QY 194 VODGQLVAL-PDKSTGLAN-----PCQFAGYTGAASEPTSV 229
DB 414 DDGGQFVVTNPNVNDGILKTAKGLDFAKQYILHVAVTNVVPEVSLTITSTATVTVDV 473
QY 230 LIINHLHIEILIDPESQVGTDRAGV-----KDVILESAITTIMDPEDSVAAV 278
DB 474 LDVNEG---PIFVPEKREVESEDFGVGQEITSYTAQEPDTFMEQKITRYI--WRDT---- 525
QY 279 DAADKVLGYRWNLGN--KGLAAAVDKDGTAFILVLRNDRNYTAPGGQFTLPGSLMF 336
DB 526 ---RVLEINPDTGAISTRAELDREDFEHVKN--STYTA----- 559
QY 337 VRNVGHMTNDAIVDTD--GSEVFEGIMDALFTGLIAHGLKASDV--NGPLINRTGSIY 393
DB 560 ---LIATDNGSPVATG-----TGTL--LILSDVNDNAPIPEPTPIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL-FSRV-----EDVLGLPQNTMKIGIM 431
DB 600 ERNPKQVINIHDADLPNTSPFTAEALTHGRVNPNTIYQNDPTQESIIILKPKMALEVG-- 657
QY 432 DEERTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMAGPMVRKGTMKSQPWILAY 491
DB 658 ---DYKINLKLMDNQNKDQVTTLEVSVCDEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSGRAQVKGKMTMTLMADVMETKIAQPRAGASTAVPSPATA 547
DB 697 ---VEAGLQIPAILGILG-----GILAL--LILILLLLFLRRRAVVKPELLP-PEDD 743
QY 548 TLHALHYH-----QVDVAAVQOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
DB 744 TRDNVYVYDEGGEGEDQDFLSQLHRLGDARPEVTRNDVAPTLMSVPRYLPRANPDEI 803
QY 594 REEVNQCOS 603
DB 804 GNFDENLKA 813

```

RESULT 9

```

US-08-732-429-2
; Sequence 2, Application US/08732429
; Patent No. 630080
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,429
; FILING DATE: herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/237,919
; FILING DATE: 3 May 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids

```

RESULT 10

```

US-09-798-267-2
; Sequence 2, Application US/09798267
; Patent No. 6406870
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael
; APPLICANT: Cepek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; INTERACTIONS WITH T LYMPHOCYTES
; FILE REFERENCE: L0560/700SERP
; CURRENT APPLICATION NUMBER: US/09/798,267
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: US 08/732,429
; PRIOR FILING DATE: 1996-11-01
; NUMBER OF SEQ ID NOS: 9

```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-798-267-2

Query Match
  3.4%; Score 128; DB 4; Length 878;
Best Local Similarity 22.0%; Pred. No. 0.004;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY 91 LPEPDDFTI-----TTSGVDABEITTTAGPQLVVPVFNARFALNAANARWGLSYDAL 141
DB 307 LPDKXMTINRNTGVISVTTGLDRESFPTY--TLVV-----QAADLQEGSL--ST 353
QY 142 YGTDVIPETDGAKEGPTYNKV--RGD-----KVAYARKFLDDSVPLSSGSGFGDATGFT 193
DB 354 TATAVITVTDNDNPPIFNPTTKGQVPENANVITLKVTDADAPNTPAWEAVYITLN 413
QY 194 VQDQQLVVAL-PDKSTGLAN-----PGQFAGYTGAAESPTSV 229
DB 414 DGGQFVVTTNPVNDGILKTAKGLDPEAKQYILHVAVTNVVPEVSITTSATVTVDV 473
QY 230 LLINHGHLHIELIDPESQVGTDRAGV-----KDVILESAITTIMDFEDSVAAV 278
DB 474 LDVNEG---PIFVPEKRVESDEFGVQGEITSYTAQEPDTFMEQKITRYI--WRDT---- 525
QY 279 DAADKVLGYRNMLGLN--KGDAAAADKDGTAFLVLRNDRNYTAPGGQFTLPGRLMF 336
DB 526 -----RNWLEINPDTGAISTRAELDREDFEHVKN--STYTA----- 559
QY 337 VRNVGHLMTNDIVDTP--GSEVFEIGMDALFTGLIAIHGLKASDV--NGPLINSRTGSIY 393
DB 560 -----LIIATDNGSPVATG-----TGTL--LILSDVNDNAPIPEPRTIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL-FSRV-----EDVLGLPQNTWKIGIM 431
DB 600 ERNPKQVINIHDADLPNTSPFTAEITHGRVNPNTIQVNDPTQESIIILKPKMALEVG-- 657
QY 432 DEERTTIVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPVRKGTMTKSQPWILAY 491
DB 658 ----DYKINLKLMDNQNKDQVTTLEVSVCDEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSGRAQVGKGMWTMTLMADWVETKIAQPRAGASTAWVPSPTAA 547
DB 697 ----VEAGLQIPAILGILG-----GILAL--LILILLLLFLRRRAVVKPELLP-PEDD 743
QY 548 TLHALVH-----QVDVAAVOOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
DB 744 TRDNVYYVDEEGGEEEDQDPLSQLHRGLDARPEVTRNDVAPTLMSVPRYLPRPANPDEI 803
QY 594 REEVNDCQS 603
DB 804 GNFDENLKA 813

RESULT 11
US-09-798-267-3
; Sequence 3, Application US/09798267
; Patent No. 6406870
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael
; APPLICANT: Cepek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; FILE REFERENCE: L0560/7008EP
; CURRENT APPLICATION NUMBER: US/09/798,267
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: US 08/732,429

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-798-267-2

Query Match
  3.4%; Score 128; DB 4; Length 878;
Best Local Similarity 22.0%; Pred. No. 0.004;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY 91 LPEPDDFTI-----TTSGVDABEITTTAGPQLVVPVFNARFALNAANARWGLSYDAL 141
DB 307 LPDKXMTINRNTGVISVTTGLDRESFPTY--TLVV-----QAADLQEGSL--ST 353
QY 142 YGTDVIPETDGAKEGPTYNKV--RGD-----KVAYARKFLDDSVPLSSGSGFGDATGFT 193
DB 354 TATAVITVTDNDNPPIFNPTTKGQVPENANVITLKVTDADAPNTPAWEAVYITLN 413
QY 194 VQDQQLVVAL-PDKSTGLAN-----PGQFAGYTGAAESPTSV 229
DB 414 DGGQFVVTTNPVNDGILKTAKGLDPEAKQYILHVAVTNVVPEVSITTSATVTVDV 473
QY 230 LLINHGHLHIELIDPESQVGTDRAGV-----KDVILESAITTIMDFEDSVAAV 278
DB 474 LDVNEG---PIFVPEKRVESDEFGVQGEITSYTAQEPDTFMEQKITRYI--WRDT---- 525
QY 279 DAADKVLGYRNMLGLN--KGDAAAADKDGTAFLVLRNDRNYTAPGGQFTLPGRLMF 336
DB 526 -----RNWLEINPDTGAISTRAELDREDFEHVKN--STYTA----- 559
QY 337 VRNVGHLMTNDIVDTP--GSEVFEIGMDALFTGLIAIHGLKASDV--NGPLINSRTGSIY 393
DB 560 -----LIIATDNGSPVATG-----TGTL--LILSDVNDNAPIPEPRTIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL-FSRV-----EDVLGLPQNTWKIGIM 431
DB 600 ERNPKQVINIHDADLPNTSPFTAEITHGRVNPNTIQVNDPTQESIIILKPKMALEVG-- 657
QY 432 DEERTTIVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPVRKGTMTKSQPWILAY 491
DB 658 ----DYKINLKLMDNQNKDQVTTLEVSVCDEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSGRAQVGKGMWTMTLMADWVETKIAQPRAGASTAWVPSPTAA 547
DB 697 ----VEAGLQIPAILGILG-----GILAL--LILILLLLFLRRRAVVKPELLP-PEDD 743
QY 548 TLHALVH-----QVDVAAVOOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
DB 744 TRDNVYYVDEEGGEEEDQDPLSQLHRGLDARPEVTRNDVAPTLMSVPRYLPRPANPDEI 803
QY 594 REEVNDCQS 603
DB 804 GNFDENLKA 813

RESULT 12
PCT-US95-05518-2
; Sequence 2, Application PC/TUS9505518
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
```

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/05518
 FILING DATE: herewith
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/237,919
 FILING DATE: 3 May 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Plumer, Elizabeth R
 REGISTRATION NUMBER: 36,637
 REFERENCE/DOCKET NUMBER: B0601/7023
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 878 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-05518-2

Query Match 3.4%; Score 128; DB 5; Length 878;
 Best Local Similarity 22.0%; Pred. No. 0.004;
 Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY	91	LPEDDDFTI-----TTSGVDAEITTTAGPOLVVFVNLNARFALNANARMSGLYDAL	141
DB	307	LPDKMFTINRTGVSIVTTGLDRESFPTY--TLVV-----QAADLQEGGL--ST	353
QY	142	YGTDVIPETDGAEGPYNKV--RGD-----KVIAARFELDSVPLSSGSGDGTGFT	193
DB	354	TATAVITVTDNDPPIFNPTTYRGQVPEANEVITTLKVTADAPNTPAWEAVYTLN	413
QY	194	VQDGQVLVAL--PDKSTGLAN-----PGQFAGTGAESPTSV	229
DB	414	DGQGFVVTNPVANDGILKTAKGLDFEAKQOYLHVAVTNVVPFVSLTSTATVTDV	473
QY	230	LLINHLHIELIDPESQVGTTRDAGV-----KDVILESAITTIMDFEDSVAAV	278
DB	474	LDVNEG--PIFVPEKRVESDFGVQBEITSYTAQEPDTFMEQKITRYI--WRDT--	525
QY	279	DAADKVLGYRNWGLN--KGPLAAAVDKGTAFRLVRLNDRNNTAPGGGQFTLPGSLMF	336
DB	526	-----RWNLINPDTGAISTRAELDREDFEHVKN--STYA-----	559
QY	337	VRNVGHMTNDAIVDTP--GSEVFGINDALFTGLIAHGLKASDV--NGPLINSRTGSIY	393
DB	560	-----LIITADNGSPVATG-----TGTL--LILSDVNDNAIPEPRITFFC	599
QY	394	IVKPK-----MHG-----PAEVAFTCEL--FSRV-----EDVLGLPQNTMKIGIM	431
DB	600	ERNPAQVINHDADLPNTSPFTFAELTHGRVPNWTIOYNDPTQESLILAPKMALEV--	657
QY	432	DEERTTVNLKACIKAAADRNVFINTGLDRTGDEIHTSMGAFMVRKMTKSPWILAY	491
DB	658	-----DYKINLKLMDNQNDQVTTLEVSVCDEG-----AAGVCR-----KAQP-----	696
QY	492	EDHNVADGL-----AAGSGRAQVCKGNTWMTLMADWVETKIAQPRAGASTANVPSPTAA	547
DB	697	-----VEAGLQPAILGILG-----GILAL--LILILLLLLFURRAVVKPELPL--PEDD	743
QY	548	TLHALHYH-----QVDVAAVQOGLAK-----PRATIEQLLTIP--LAKELAWAPDEI	593
DB	744	TRDNVYYVDEGGGEEDDFLSQLHGLDARPEVTRNDVAPTLMSVPRVLPAPNPDEI	803
QY	594	REEVNNCOS 603	
DB	804	GNFIDENLKA 813	

RESULT 13

US-09-105-537-6
 ; Sequence 6, Application US/09105537A

; Patent No. 6265202
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600.438US1
 ; CURRENT APPLICATION NUMBER: US/09/105,537A
 ; CURRENT FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 11877
 ; TYPE: PRT
 ; ORGANISM: Streptomyces venezuelae
 US-09-105-537-6

Query Match 3.4%; Score 128; DB 3; Length 11877;

Best Local Similarity 20.7%; Pred. No. 0.34;
 Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;

QY	9	LRIARVLDFVNEALPGTDIDPDSFWAGVKKVADLTPOQALLNADELOAQIDKWHR	68
DB	10193	LRDAGVL-----DTVRLTGTIEPEFGSGSGDGAADPGAEPEASI---DLDLDAEA---LI	10241
QY	69	RVIEPID-----MDAYRQLTEITGYLLPPDPDDFTITTSVDAEITTTAGPOLVVP	119
DB	10242	RNALGFRMTWSSNEQLVDALRASLKENEELKES-----RRADRRQEPMAIVG	10291
QY	120	VLNARFALNANARMSGLYDAL--YGTDV---IPETDGAEGKGTYNKVRGDKVIAYARK--	173
DB	10292	-MSCRFAGGIRSP--DLMDAVAAGKDLVSEVPEERGWDIDSLYDPVGRKGTTVYRNAA	10348
QY	174	FLDDSVPLSSGSGF---DATGFTVQDQVLVALPKSTGLANPGQAPAGYGAASPTSV	229
DB	10349	FLDDAAGDDAAFFGISPREALAMPQOQLLEA-----SWEVFERAGIDFASV	10396
QY	230	LLINHLHI-----ELIDPESQVGTTRDAGVKDVILESAITTIMDFEDSVAAVDA	280
DB	10397	RGTDVGVVGGCYQDYAPDIRVAPEGTGGYV--VTGNSSAVASGRIAYSLGLEGPATVDT	10455
QY	281	A---DKVLGYRNWGLNKGDLAAAVDKGTAFRLVRLNDRNNTAPGGGQFTLPGESLMFV	337
DB	10456	ACSSSLVALHIALKGLRNGDCSTAL--VGGVAVL-----ATPGAFIEFS	10497
QY	338	-----RVNGHMTNDAIVDTPDSEV--FEGIMDALFTG--LIAIHGLKASDVNGPL	384
DB	10498	SQQAAVADGRTKGPASAAADGLAWGEGVAVLLERLSDAARRKGRHVLA-------VRGSA	10550
QY	385	INSRTGSIYVKKPKXKHGAEVAFCELFSEVEDVGLPQNTMKIGIMDEERTTVNLKAC	444
DB	10551	INQDGASNLTPA--HGFSQ-----OHLIROLALD--ARLT-----	10582
QY	445	IKAAADRNVFINTGLDRTGDEIHT-----SMEAGPMVRKMTKSGQ-----	485
DB	10583	---SSDSDVVEGCHGTGLDGFIEAQUALATYQGCRAPGQLRLGLTKSNIGHTQAASGV	10639
QY	486	-----PWILAYED--HNVD--AGLA-----	508
DB	10640	AGVIRKQVQALRHGVLPKTLHVDEPTDQVDWAGSAGSVELLTEAVDWPFRGLRRAGVSAGF	10699
QY	509	QVKGKMTWMTLMADWVETKIAQPRAGASTA--W--VPSPTAATL-----HALHYHOV	557
DB	10700	VGCTNAHVLEAPAEVESPAPVPGGVVFWPVSAKTSAAALDAQIQLAAAYAEEDTDV	10759
QY	558	DVAAVQOGL-----AGRRATIEQLTIP--LAKELAWAPDEI-----	593

Db 10760 DPAVARALVDSRTAMEHRAVAVGDSREALRDLRMPEGLVRGTVDPRGAVFVFGQGT 10819
QY 594 -----REEVDNN-----CQSILGYVVRVWDOGV--GCSKVPDIHDVAMEDRATL 636
Db 10820 QWAGYGNELDDSSPEFAANAAECETALSPYVDHSLAVVRQAPSATILDRVDVQV--VTF 10878
QY 637 RISSQLLANWLRHGVTISADVRASLERMAPLVDQRONAGDVAVRMAPNFDDSS 688
Db 10879 AVMSLAKVWQHGHITPEAVIGHISQGEIA-----AAYVAGALTLLDDA 10920
RESULT 14
US-09-252-991A-21798
; Sequence 21798, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21798
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21798
Query Match 3.3%; Score 127; DB 4; Length 1650;
Best Local Similarity 20.2%; Pred. No. 0.015;
Matches 177; Conservative 110; Mismatches 349; Indels 242; Gaps 37;
QY 7 GNLRIARVLVDFVNNALPCTDIDPDSFWAGV-----DKVVADLTPOQALLNARDEL 59
Db 382 GNLAAVAVEDVOPDQAQRAASTDQATLVVQVAAPOVETDAALAD--OPALALHQAIDR 438
QY 60 QAQIDKWHRRRRIEIPIDMAYROFLTEIGYLLPEPDDFTITTSQVDAE-ITTAGPQLVY 118
Db 439 QAQVAPGDFAAVGGVDAPC-RQFQFAVA---CQQAATVVQVGGAAQORLLATEGAATV 494
QY 119 PVLNARFALNARNWGLSYDALYGTDPITDQ-REKFTYKVRGDKVIA-----169
Db 495 EV-----GAARDVQ-HALAQOATLGVQVORASGDAQG-----VAGEOALAAVVOGGAG 541
QY 170 -----YARKFLD---DSVPLSSGSPG---DATGFTVQ-----DGQLVVAL 203
Db 542 ERQAGAGNGAFVGVVQRAVDAQVDAAGAGGALGAVVQASGRDVQAGGADHPGQAVV-- 599
QY 204 PDKSTGLANPGQAGVTGAESTSVLLINHG-----LH-----237
Db 600 -----QGLGDAQGQAG--GADQAPAAVVQAGGEGEAGVAGDFPAGAVVHRAELAQOQRA 653
QY 238 -----IEILIDPESQVGTDRAGVKDVILSAITTIMDFEDSVAAVDAADKVLGYRNWGL 293
Db 654 RGDQAAVAVDQRAAIEVEGDGAFADODASAL--VFAFQVGEQALGADPSLLAVVQPG 711
QY 294 NKGD-----LAAAVDKDGTAFRLVNRDRNYTA-----321
Db 712 HOGDAGVAADAATAVAVVQAGADITHRTIGADHAGTAVWEAGALQRHAGIAEQAPALVQR 771
QY 322 --PGGQFTLPGRLSFLNVRVGHMTNDALVTDGSEVEGIMDALFTGLIAHGL---- 375
Db 772 GLAQQRQTGAGGPRATVQVQARGQAAPADQRAALVVQCHAAHAQAVLAVEPTAVAV 831
QY 376 -KASVDNGLPLNRSITGSIVIVKPKMHGPAEVAFTCELSRVEDVLG-----420
Db 832 EQFAVQAQAVAPGQHPGLGLVQALHGEAQAAVADDLAAVQLLAGVHGLRGAGNLAG 891

QY 421 ----LPQ-----NTMKIGIMDEERRITTVNLKA-----CIKAAADRVRVFIINTFLDR 462
Db 892 AVVDLPRLDSDAALRGDQPLAVVDRVGRDLQGLFADFQFATLLGQAARKRLQVALGGTTP 951
QY 463 TG--DEIHTSWEAGPWRKGTMSQPMILAYEDHNVDAGLAA-----GFSGRAGVG 511
Db 952 SGVADGVRCEVQAAIAEQALTLV-----VAQAGDLHAHASLAAGAAVAVWEGASRGAGAG 1007
QY 512 KG-MWMTMELMADVVETKIAQPRAGASTAWVPSPATAATLHALHYHQVDVAAVQOGLAGKR 570
Db 1008 VGHQHAAIVVOAGAVDVQLAARQAQAGAVVOLPHRHARALALAEQVAFIAVQO-----1061
QY 571 RATIEQLLTIPLA-KELAWAFDEIREEVNDCQSIGLYVVRWVDOGVGCSKVPDIHDVAL 629
Db 1062 -AAEADLOAAPAQAORALAAAVVQAARPV-----EAVGCGEQAE-----AV 1100
QY 630 MEDRATLRISSQLLANWLRHGVTISADVRASLERMAPLVDQRONAGDVAVRMAPNF--DS 688
Db 1101 VD--GTAGDHOQVADQALAAIVAEATD--AGLDAL-----AGDFAVPPVDORLGADA 1148
QY 689 TAFLLAAQELIISGACQPNQNGYTEPILHRRRRREFKARAAE 726
Db 1149 GGLLAAEQALLA-----VVERRADLQVVQAD 1175
RESULT 15
US-09-320-878-2
; Sequence 2, Application US/09320878A
; Patent No. 6117859
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1998-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1998-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-2
Query Match 3.3%; Score 126.5; DB 3; Length 3739;
Best Local Similarity 21.4%; Pred. No. 0.066;
Matches 166; Conservative 87; Mismatches 297; Indels 227; Gaps 45;
QY 24 LPGTDIDPDSFW--AGVDKVVADLTPOQALLNARDELQAQIDKWHRRRIEIPIDMAY 80
Db 1551 LPFGVAGFEDLWRLVAGEDAISGF-PQDRG-----WDVEGLYDP-DPDAS 1594
QY 81 RQFITEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPVNLNARFALNARNWGLSYDA 140
Db 1595 QRTVCRAGGFLDEAGEPDADPFGISPREALAMDQ-----QRLLETS---WEAVEDA 1644
QY 141 LYGVTDVIPETDGAEGKPTYNKVRGDKVIAYA-----RKFLDSDSVLSSGSGDGTGCT 193

Db 1645 --GID-----PT--SLOGQVGVFAGTNGPHYEPPLRNRTAEDLEGYVGTGNAAS 1689
 QY 194 VODGQVVALPDKSTGLANPCQFAGYTGAAESPTSVLLINHGHLHIEILIDPESQVGTDDR 253
 Db 1690 IMSGRVSTL-----GLEGPV-----TVDTACSSSLV--ALHLAYQALRKCECGLALA 1736
 QY 254 AGVKDVLILESITIMDP-----ED--SVAAYDAADKVLGYRNWGLNKGDLAAAVDK 304
 Db 1737 GG--VTWMSPTTFVEPSRQGLAEDGRSKAFASADG-FCPAEGVGMILLVERLSDARR 1792
 QY 305 DGTAFIRV-----LNRD--RNYTAPGGQQTLCRSIMFVRNVGHLMTHDAIVDTDGSE 356
 Db 1793 NGRVLAHVVRGSANQDASNGLTAPNG-----PSQORVIRRALADARLTATDUD 1842
 QY 357 VFEIGMDALFTG-----LIAIHGLKASDVNGPL-----INSR-----TGS 391
 Db 1843 ---VVEAHGTGTRLGDPLEAQUALIATVG-QGRDTEQPLRLGSLKSNIGHTQAAAAGVSGI 1897
 QY 392 IYIVKPKVHGPAEVAFTCELSERVEDVLGLPQNTMKI--GIMDEERTTVNLKACIKAAA 449
 Db 1898 IKMVOAMRHGVLKPTLHVD---RPSDQIDWSAGTVELLTEAMDWPKOEGGLR----- 1947
 QY 450 DRVVFINTGFLDRDGTDEHTSMEAGPMVRKGTMSQ-----PWILAYEDHNVDAAGLAAG 503
 Db 1948 -RAVVSFTGI---SGTNAHIVLEAPVDEADAPADEPSVGVVPWL-----VSAKTPAA 1996
 QY 504 PSGRAQVKGKMWMTTELMAVMETKIAOPRAGASTAWVPSPTAATLH---ALHYHQVDVA 560
 Db 1997 LD--AQIGR-----LAAFASQGRDADP--GAVARVLAGGQAQFEHRAVALGTGQDDLA 2047
 QY 561 ---AVQOGL-----AGKERATI-----EOLLTIPLAKELAWAPDEIREVDN 599
 Db 2048 AALAAPEGLVRGVASGVGRVAFVFPQGTQWAGMGAEILDV--SKEFAAAVAE----- 2098
 QY 600 NCQSILGYVVRW-----VDQGVGCSKVPDIHDVALMEDRATLRISSOLLANWLRHGVITS 654
 Db 2099 -CEAALAFYDWSLEAVVRQAPG--APTLEKVDVQOP-VTFVAVVSLAKVQHHGVTPQ 2153
 QY 655 ADVRASLERMAPLYDRQAGDVAYRPMAPNFDDSI-----AFLAAQELILS 700
 Db 2154 AVVGHSQGEIA-----AAYVAGALSUDDAARVVTLRSKISGAHLAGQGGNLS 2200

Search completed: November 21, 2003, 16:11:36
 Job time : 18.6054 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 / Search time 44.8027 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-2

Perfect score: 3810

Sequence: 1 TDRVSGNLRIRARVLYDFVN.....KARAAEKAPSDRAGDAAR 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 segs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	100.0	741	AAE05665	Mycobacterium tube
2	3810	100.0	1172	AAU01900	M. tuberculosis tb
3	3804	99.8	2502	AAE00117	M. tuberculosis po
4	2289	60.1	739	AAE077501	Malate synthase.
5	2289	60.1	739	AAE03234	C glutamicum prote
6	2275	59.7	739	AAE79539	Corynebacterium gl
7	2253	59.1	712	AAE79540	Corynebacterium gl
8	991.5	26.0	1190	ABG18357	Novel human diagno
9	741.5	19.5	830	ABG25453	Novel human diagno

10	594.5	15.6	504	22	ABG18358	Novel human diagno
11	585.5	15.4	1641	22	ABG18356	Novel human diagno
12	433.5	11.4	441	22	ABG18355	Novel human diagno
13	420	11.0	130	23	ABP03007	Human ORFX protein
14	189.5	5.0	533	13	AAE037585	E.coli malate synt
15	189.5	5.0	533	24	AAO16702	Escherichia coli a
16	135	3.5	3816	21	AAE92708	S. antibioticus 8,
17	134.5	3.5	1440	23	ABBS4801	Lactococcus lactis
18	130	3.4	1624	23	ABU05888	M. tuberculosis an
19	128.5	3.4	823	22	ABG09169	Novel human diagno
20	128.5	3.4	823	22	ABG26009	Novel human diagno
21	128	3.4	878	15	AAE55060	Sequence of human
22	128	3.4	878	16	AAE54847	Human E-cadherin p
23	128	3.4	878	23	ABB1475	Human E-cadherin p
24	128	3.4	12199	21	AAE77180	S. venezuelae pik
25	127.5	3.3	5701	22	ABG51536	Human liver peptid
26	127.5	3.3	5701	22	ABG36684	Peptide #4190 enco
27	127.5	3.3	5701	22	ABB22021	Protein #4020 enco
28	127.5	3.3	5701	22	AAE57448	Human brain expres
29	127.5	3.3	5701	22	AAE59843	Human bone marrow
30	127.5	3.3	5701	22	AAE05328	Peptide #4010 enco
31	127.5	3.3	26926	22	AAU05396	Human titin (conne
32	127.5	3.3	31267	24	ABG74786	Human RSL1 protei
33	126.5	3.3	3739	21	AAE18638	Amino acid sequenc
34	126.5	3.3	3739	21	AAE77193	S. venezuelae macr
35	126.5	3.3	3739	21	AAE77201	S. venezuelae pik
36	126.5	3.3	3739	21	AAE77202	Marbonolide syntha
37	126.5	3.3	3739	24	ABG71662	S. venezuelae narb
38	125.5	3.3	2039	19	AAE56322	Haemophilus paraga
39	125	3.3	918	20	AAE98422	Chlamydia pneumoni
40	124	3.3	1346	21	AAE77195	S. venezuelae macr
41	124	3.3	4999	23	AAO22158	Ramoplanin biosynt
42	124	3.3	15281	15	AAE44929	T. niveum Cyclospo
43	123	3.2	1346	21	AAE18640	Amino acid sequenc
44	123	3.2	1346	21	AAE7204	Marbonolide syntha
45	123	3.2	1346	24	ABG71664	S. venezuelae narb

ALIGNMENTS

RESULT 1

AAE05665

ID AAE05665 standard; peptide; 741 AA.

XX AC AAE05665;

XX DT 24-SEP-2001 (first entry)

XX DE Mycobacterium tuberculosis 88 kDa secreted protein.

XX KW Mycobacterial disease; infection; 88-kDa protein; tuberculosis; TB;

XX OS Mycobacterium tuberculosis.

XX PN US6245331-B1.

XX PD 12-JUN-2001.

XX PF 31-DEC-1997; 97US-0001984.

XX PR 02-JAN-1997; 97US-0034003.

XX PA (UNIV-) UNIV NEW YORK MEDICAL CENT.

XX PA (COLS) UNIV COLORADO STATE.

XX PI Laal S, Zolla-Pazner S, Belisle JT;

XX DR WPI; 2001-424324/45.

XX PT Detecting a mycobacterial disease (tuberculosis) in individuals

XX PT comprise assaying a biological sample for the presence of

PT anti-Mycobacterium tuberculosis antibodies or M. tuberculosis
XX antigen-antibody complex
PS Claim 14; Column 121; 96pp; English.

The present invention relates to a method for early detection of active mycobacterial disease or infection comprising assaying a biological fluid sample for the presence of early antibodies specific for an 88-kDa Mycobacterium tuberculosis (Mt) protein or immune complexes consisting of an 88-kDa M. tuberculosis protein antigen complexed with an antibody specific for the antigen. The method is useful for the early and rapid detection of mycobacterial disease, particularly tuberculosis, in individuals at heightened risk of developing tuberculosis. This method includes human immunodeficiency virus (HIV)-infected subjects or other immunocompromised individuals. The method is a rapid and inexpensive screening procedure for detecting mycobacterial disease. The present sequence is a Mt 88 kDa secreted protein early antigen, used in the exemplification of the invention.

XX Sequence 741 AA;

Query Match 100.0%; Score 3810; DB 22; Length 741;
Best Local Similarity 100.0%; Pred. No. 4,7e-316;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVQKVVADLTQNOALLNARDELQ 60
DB 2 TDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVQKVVADLTQNOALLNARDELQ 61
QY 61 AQIDKWHRRVIEPIDMDAYQFITEIGYLLPEPDDFTITTSQVDAITTAGQLVVPV 120
DB 62 AQIDKWHRRVIEPIDMDAYQFITEIGYLLPEPDDFTITTSQVDAITTAGQLVVPV 121
QY 121 LNARFALNAARWGSGLYDALYGVTDVTPETDGAEGKPTYNKVRGDKVIAYARFELDDSV 180
DB 122 LNARFALNAARWGSGLYDALYGVTDVTPETDGAEGKPTYNKVRGDKVIAYARFELDDSV 181
QY 181 LSSGSGFGDGTFTVDGQVLVALPKSTGLANPQFAGYTGAAESPTSVLLINHLHIEI 240
DB 182 LSSGSGFGDGTFTVDGQVLVALPKSTGLANPQFAGYTGAAESPTSVLLINHLHIEI 241
QY 241 LIDPESQVGTTRAGVQKVVILESAITTIMDEDSVAADKVLGYRNWGLNKGDLAA 300
DB 242 LIDPESQVGTTRAGVQKVVILESAITTIMDEDSVAADKVLGYRNWGLNKGDLAA 301
QY 301 AVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFRVNVGHLMTNDALVDTGSEVFEG 360
DB 302 AVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFRVNVGHLMTNDALVDTGSEVFEG 361
QY 361 IMDALFTGLIAIHGLKASDVNGPLINSGTGIYIVKPMHGPFAEVAFTCELFSEVDVLG 420
DB 362 IMDALFTGLIAIHGLKASDVNGPLINSGTGIYIVKPMHGPFAEVAFTCELFSEVDVLG 421
QY 421 LPONTMKIGIMDEERTTVNLKACIKAAADRVWINTGFLDRTGDEIHTSMEAGPMVRKG 480
DB 422 LPONTMKIGIMDEERTTVNLKACIKAAADRVWINTGFLDRTGDEIHTSMEAGPMVRKG 481
QY 481 TMKSQFWILAYEDHNVDAGLAGSFGRAQVKGKWWMTLMADVVEVKIAPRAGASTAW 540
DB 482 TMKSQFWILAYEDHNVDAGLAGSFGRAQVKGKWWMTLMADVVEVKIAPRAGASTAW 541
QY 541 VPSPTAATLHALHYHVDVAVVQOGLACKRATTEQLLTIFLAKELAWAPEIRIEVDNN 600
DB 542 VPSPTAATLHALHYHVDVAVVQOGLACKRATTEQLLTIFLAKELAWAPEIRIEVDNN 601
QY 601 COSILGYVVRWVDQGVGSKVPDIHDVALMEDRATLRISQQLLANWLEHGVITSADYRAS 660
DB 602 COSILGYVVRWVDQGVGSKVPDIHDVALMEDRATLRISQQLLANWLEHGVITSADYRAS 661
QY 661 LERMAPLVDQRNAGVAVRPAPNFDSDIAFLAAQELTSLGAQOPNGVTEPILHRRREF 720
DB 662 LERMAPLVDQRNAGVAVRPAPNFDSDIAFLAAQELTSLGAQOPNGVTEPILHRRREF 721

QY 721 KARAAEKPAAPSDRAGDDAAR 740
DB 722 KARAAEKPAAPSDRAGDDAAR 741

RESULT 2

AAU01900

ID AAU01900 standard; Protein; 1172 AA.

XX AAU01900;

DT 29-AUG-2001 (first entry)

XX M. tuberculosis Tbf14 fusion protein.

XX Tbf14; antigen; vaccine; tuberculosis; AIDS; His tag; MtB81; Mo2;

XX acquired immunodeficiency disease.

XX Synthetic.

OS Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT Binding-site 3..8

FT /label= Histidine_tag

FT /note= "Nickel chelating region used for purifying

FT the fusion protein"

FT Region 9..749

FT /label= MtB81_region

FT Misc-difference 49

FT /note= "Encoded by TC"

FT Region 750..1172

FT /label= Mo2_region

XX WO200124820-A1.

XX 12-APR-2001.

XX 10-OCT-2000; 2000WO-US28095.

XX 07-OCT-1999; 99US-0158338.

XX 07-OCT-1999; 99US-0158425.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Iodes ML;

XX WPI; 2001-290576/30.

XX N-PSDB; AA303794.

XX Vaccinating against Mycobacteria infections in mammals using fusion

XX proteins comprising combinations of heterologous antigens -

XX Claim 3; Fig 3; 168pp; English.

XX The sequence represents Mycobacterium fusion protein antigen Tbf14
XX consisting of a His tag for purification, antigen MtB81 and antigen Mo2.
XX Compositions comprising at least 2 heterologous antigens as a fusion
XX protein, and vectors expressing the fusion proteins are used as vaccines
XX to prophylactically immunise mammals (especially humans) against
XX infection by Mycobacteria. The compositions contain at least 2
XX heterologous antigens that increase the serological sensitivity of
XX individuals infected with tuberculosis, a disease frequently affecting
XX patients with acquired immunodeficiency disease, AIDS.

XX Sequence 1172 AA;

Query Match 100.0%; Score 3810; DB 22; Length 1172;

Best Local Similarity 100.0%; Pred. No. 9.3e-316;

Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVQKVVADLTQNOALLNARDELQ 60
DB 9 TDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVQKVVADLTQNOALLNARDELQ 68

FT Misc-difference 1035 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1082 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1160 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1161 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1164 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1187 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1188 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1211 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1253 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1344 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1348 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1386 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1527 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1579 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1582 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1616 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1628 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1643 /note= "Encoded by TGA stop codon"
 FT Misc-difference 1673 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1683 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1685 /note= "Encoded by TAA stop codon"
 FT Misc-difference 1691...2438 /note= "Encoded by TAA stop codon"
 FT Protein /label= Mtb-81 polypeptide
 FT Misc-difference 2439 /note= "Encoded by TAG stop codon"
 FT Misc-difference 2440 /note= "Encoded by TGA stop codon"
 FT Misc-difference 2464 /note= "Encoded by TAA stop codon"
 FT Misc-difference 2471 /note= "Encoded by TGA stop codon"
 FT Misc-difference 2478 /note= "Encoded by TGA stop codon"
 FT Misc-difference 2489 /note= "Encoded by TAA stop codon"
 FT WO200055194-A2.
 XX 21-SEP-2000.
 XX 17-MAR-2000; 2000WO-US07196.
 XX 18-MAR-1999; 99US-0272975.
 XX (CORI-) CORIXA CORP.
 XX Hendrickson RC, Lodes MJ, Houghton RL;
 XX WPI; 2000-638180/61.
 XX N-PSDB; AAA53971.
 XX Novel Mycobacterium tuberculosis polypeptide comprising an immunogenic

PT portion of M. tuberculosis antigens Mtb-81 and Mtb-67.2, useful for
 PT diagnosis, treatment and monitoring therapy of tuberculosis
 XX Claim 1: Fig 1a-f; 9lpp; English.
 CC Polypeptides comprising an immunogenic portion of Mycobacterium
 CC tuberculosis antigens Mtb-81 or Mtb-67.2 or variants that differ
 CC by substitutions, additions, insertions and/or deletions but which
 CC still react with antigen specific antisera or T-cells are described.
 CC Also described are polynucleotides encoding the polypeptides. The
 CC polypeptides, expression vectors expressing the polypeptides or
 CC comprising an antisense polynucleotide, or an antigen presenting
 CC cell comprising a sequence encoding the polypeptides are useful for
 CC determining the presence or absence of M. tuberculosis in whole blood,
 CC serum, sputum, plasma, saliva, cerebrospinal fluid or urine in a
 CC patient infected with human immunodeficiency virus (HIV).
 XX Sequence 2502 AA;
 SQ Query Match 99.8%; Score 3804; DB 21; Length 2502;
 Best Local Similarity 99.7%; Pred. No. 9.3e-315;
 Matches 738; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TDRSVGNLRIRARVLYDFVNNALPGTDIDPDSFWAGVKVADLTPOQALLNARDLQ 60
 Db 1699 TDRSVGNLRIRARVLYDFVNNALPGTDIDPDSFWAGVKVADLTPOQALLNARDLQ 1758
 QY 61 AOIDKWHRRVTEPIDMDAYRQFLTEIGVLLPEPDDFTITTSVDAEITTTAGPOLVVPV 120
 Db 1759 AOIDKWHRRVTEPIDMDAYRQFLTEIGVLLPEPDDFTITTSVDAEITTTAGPOLVVPV 1818
 QY 121 LNARFALNAANARWGLYDALYGTDTVIBETDGAEGKPTYNKVRGDKVAYAKPFDDSDVP 180
 Db 1819 LNARFALNAANARWGLYDALYGTDTVIBETDGAEGKPTYNKVRGDKVAYAKPFDDSDVP 1878
 QY 181 LSSGSPGDATGTTVDGQGLVVALPDKSTGLANPGFAGVTGAESPTSVLLINHLHIEI 240
 Db 1879 LSSGSPGDATGTTVDGQGLVVALPDKSTGLANPGFAGVTGAESPTSVLLINHLHIEI 1938
 QY 241 LIDPESQVCTTDRAGVKDVIRESAITTIMDFDSVAADAAKVLGYRNWGLNKGDLAA 300
 Db 1939 LIDPESQVCTTDRAGVKDVIRESAITTIMDFDSVAADAAKVLGYRNWGLNKGDLAA 1998
 QY 301 AYDKDGTAFRLVNRDRNYTAPGGQFTLPGRSLFVRNVGHLMTNDAIVDTGSEVFEG 360
 Db 1999 AYDKDGTAFRLVNRDRNYTAPGGQFTLPGRSLFVRNVGHLMTNDAIVDTGSEVFEG 2058
 QY 361 IMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFACTELPSRVEDVLG 420
 Db 2059 IMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFACTELPSRVEDVLG 2118
 QY 421 LPQNTWKIGIMDEERTTNLXACIKAAADRVVFTNGTFLDRTGDEIHTSMGAPWRKG 480
 Db 2119 LPQNTWKIGIMDEERTTNLXACIKAAADRVVFTNGTFLDRTGDEIHTSMGAPWRKG 2178
 QY 481 TMKSQPWILAYEDHNVDAGLAAGFSGRAQVKGKMTMTLMADVMVETKIAQPRAGASTAW 540
 Db 2179 TMKSQPWILAYEDHNVDAGLAAGFSGRAQVKGKMTMTLMADVMVETKIAQPRAGASTAW 2238
 QY 541 VPSPTAATLHALHYHQVDVAAVQOGLAGKRATITBOLLTIPLAKELANAPDSIREVDNN 600
 Db 2239 VPSPTAATLHALHYHQVDVAAVQOGLAGKRATITBOLLTIPLAKELANAPDSIREVDNN 2298
 QY 601 CQSILGVVVRWVDQGVGCKVPDIHDVALMEDRATILRISQLLANLWLRHGVITTSADVRAS 660
 Db 2299 CQSILGVVVRWVDQGVGCKVPDIHDVALMEDRATILRISQLLANLWLRHGVITTSADVRAS 2358
 QY 661 LERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGACQPNGYTEPILHRRREF 720
 Db 2359 LERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGACQPNGYTEPILHRRREF 2418
 QY 721 KAPAAEKPAKPSDRAGDDAAR 740

2419 KARAABKAPSPDRAGDDAAR 2438

RESULT 4.
AAR77501
ID AAR77501 standard; Protein; 739 AA.
XX
AC AAR77501;
XX
DT 04-JUN-1996 (first entry)
XX
DE Malate synthase.
XX
KW aceB gene; malate synthase; Corynebacterium; regulatory region;
KW protein synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN DE4440118-Cl.
XX
PD 09-NOV-1995.
XX
PF 11-NOV-1994; 94DE-4440118.
XX
PR 11-NOV-1994; 94DE-4440118.
XX
(KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
XX
PA Eikmanns B, Reinscheid D, Sähm H;
XX
WP1: 1995-374802/49.
DR N-PSDB; AAT05501.
XX
PT DNA contg. upstream regulatory region from malate synthase gene of
PT corynebacterium - used for regulated synthesis of protein in
PT coryneform(s), regardless of nutrient medium compsn..
XX
PS Example 3; Page 7-8; 12pp; German.
XX
CC AAR77501 is encoded by an approx. 3 kb BfrI-PvuI fragment contg. the
CC aceB gene from Corynebacterium glutamicum (see AAT05501). Nucleotides 1
CC to 574 can be isolated and used to regulate the expression of a
CC protein-encoding structural gene placed downstream of it, after
CC incorporation into a vector and transfer into a Corynebacterium.
CC Induction of protein expression occurs even when the culture medium
CC contains carbon sources other than acetate.
XX
SQ Sequence 739 AA;

Query Match 60.1%; Score 2289; DB 16; Length 739;
Best Local Similarity 61.3%; Pred. No. 3.1e-186;
Matches 444; Conservative 90; Mismatches 184; Indels 6;- Gaps 4

QY 1 TDSVSVGLRVARVLVDYFNNEALPGTDDDPDSFWAGVDKVVADLTFQCALINARDELQ 60
DB 20 TERVDAGGNQAKVLVDYDEVTEAVLPRVGVDAEKFWSGFAAARDLTPRNRELLARDELQ 79
QY 61 AGIDKWHRRVRIEPIIDMAYRQFLTEIGYLLPEPDDFTTITSGVDAEITTTAGPOLVVPV 120
DB 80 MLIDDYHRNN-SGTDQAEYDFLKEIGLVEPEAAEIRTONVDTEISSTAGPOLVVPV 138
QY 121 INARFALNAANRWGSLYDALYGTDTVPETDGAEEKGPTYNKKRGDKVIAYARKFLDSDVP 180
DB 139 LNARFALNAANRWGSLYDALYGTNAIPETDGAEEKGYENPVRGQKVIEMWGREFLDSDVP 198
QY 181 LSSGFGDAGTGTVDGQGVVALPKSTGLANPQFAGYTGAAESPTSVLLINHGHLHEI 240
DB 199 LDGASHADVKNYITDGLAAHIGDSVYLRKRESYKGTGNFLDPEAILLETNGHLHEL 258
QY 241 LIDPESQVGTTRAGVKVILLESATITMDFPESVAADVAADKVLGYRNLWGLNKGDAA 300
DB 259 QDQVPHPTGKADKTKGLKIVLESATITMDFPESVAADVAADKTKLGYSNWGLNTGELKE 318

QY	301	AVDXDGTAFVLNRDRNYTAPGGQFTLPGRSLMFVRNVGHLMNTDAIVDTGDSGEVFEG	366
Db	319	EMSKNGRIFTBELAKDRVYIGRNGTELVHGRSLLFVRNVGHLMQNP SIL-IDGEEIPEG	377
QY	361	IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMHPGPAEVAFTCELSRVEDVLG	420
Db	378	IMDAVLTVCAIPGLAPQN---KMRNSRKSGSIYIVKPKQHPGPEEVAFTNELFGRVEDLLD	434
QY	421	LPONTMKIGIMDEBERTTVNLKACIAAADRVVFINTGFLORTGDEIHTSMGAPMVRKG	480
Db	435	LPRLTKVGVNDEERTSVNLDASIMEVADRLAFINTGFLORTGDEIHTSMGAGMVRKA	494
QY	481	TMKSQPMILAYEDHNVDAGLAAGTSGRAQVGKGMWMTDELVADMVETKIAQPRAGASTAW	540
Db	495	DMQAPKPKQAVENNVDAGIOGGLPGKAQIGKGMWMTDELVADMVETKIAQPRAGASTAW	554
QY	541	VPSPTAATLHALHYHQVDVAAVQOGL-AGKRRTIEQLTTIPLAKELAWAPDEIEEVDN	599
Db	555	VPSPTGATLHATHYELVDFVKVQDELRAAGRRDSLRLNLTITPTAPNTNWSBEKKEEDN	614
QY	600	NCOSILGVVVRWVQGVCSKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADEVRA	659
Db	615	NCOSILGVVVRWVSHGVCSKVPDIHDIDLMDRATLRISQMLANWLRLHGVITSADEVRA	674
QY	660	SLERMAPLVRQNGADVAYRPMWAFNFDSDIAFLAAQELILSGAQOPNGVTEPILHRRRE	719
Db	675	SLERMVAVVDKQNGADRAYRPMWAFNFDSDIAFLAAQELILSGAQOPNGVTEPILHRRRE	734
QY	720	PKAR 723	
Db	735	PKAK 738	
RESULT 5			
AAG93234			
ID	AAG93234 standard; Protein; 739 AA.		
XX	AC	AAG93234;	
XX	DT	26-SEP-2001 (first entry)	
XX	DE	C glutamicum protein fragment SEQ ID NO: 6988.	
XX	KW	Coryneform bacterium; amino acid synthesis; vitamin; saccharide;	
XX	KW	organic acid synthesis.	
XX	OS	Corynebacterium glutamicum.	
XX	PN	EP1108790-A2.	
XX	PD	20-JUN-2001.	
XX	PF	18-DEC-2000; 2000EP-0127688.	
XX	PR	16-DEC-1999; 99JP-0377484.	
PR	07-APR-2000; 2000JP-0159162.		
PR	03-AUG-2000; 2000JP-0280988.		
XX	PA	(KYOW) KYOWA HAKKO KOGYO KK.	
XX	PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;	
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;		
XX	DR	WPI; 2001-376931/40.	
DR	N-PSDB; AAH68453.		
XX	Novel polynucleotides derived from Coryneform bacteria, for identifying		
PT	mutation point of a gene, measuring expression of a gene, analysing		
PT	expression profile or pattern of a gene and identifying homologous gene		
XX	Claim 29; SEQ ID NO: 6988; 246pp + Sequence Listing; English.		

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

Claim 29; SEQ ID NO: 6988; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.

CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

AX	Sequence	739 AA:
Qy	Query Match	60.1%; Score 2289; DB 22; Length 739;
Db	Best Local Similarity	61.3%; Pred.No.3.1e-186;
Matches	444; Conservative	90; Mismatches 184; Indels 6; Gaps 4;
Qy	1	TDVSVGNLRTARVLVDFVNNEALPGTDIDPDSFWAGVGVKVVADLTTPQNOALLNARDELQ 60
Db	20	TEVVDAGQWAKVLVDFVTEAVLPRVGVDAEFKWSGFAAIARDLTPRNEELLARRDELQ 79
Qy	61	AGIDKWHRRRVIEPIDMDAVRQFTEIGYLLPPDDFTTITTSGVDAEITTTAGPOLVVPV 120
Db	80	MLIDDYHRNN-SGTIDQAEVDFLKEIGYLVEEPEAAEINTQVDETSSTAGPOLVVPV 138
Qy	121	LNARFALNAANRWGSLYDALYGTVDVTPETDGAEGKPTYNKVRGDKVAYARKFIDDSVP 188
Db	139	LNARFALNAANRWGSLYDALYGTNAIPETDGAEGKKEYNPRGVQKVIEMGEEFLDSVVP 198
Qy	181	LSGSFGDATGFTVQDQGVVALPKDSTGLANPQFAGYTGAAESTSVLLINHGHIET 240
Db	199	LDGASHADVKEKYNITDGLAAHIGDSVYRLKNRESYRGFTGNFLDPEAILLTNGLHIEL 258
Qy	241	LIDPESQVTTDRAGVKDVILESATITIMDFEDSVAADAAKVLGYRNWGLNKGDLAA 300
Db	259	QIDPVHPTGKADTKGLDVILESATITIMDFEDSVAADAEKTLGYSNWFGLNTGELKE 318
Qy	301	AVDKGTAPLVLNDRNYTAPGGQFTLPGRSIMFVRNVGHLMTNDIAIVDTDGESEVEG 360
Db	319	EMS KGRIFPTRELNDRVYIIGKNETELVJHGRSLLFVRNVGHLMQNP SIL-IDGSEIFE 377
Qy	361	IMDALFTGLIATHGLKASDVNGPLINSRTGSIIYVFKPMHGPAAEVAFTCELESFVEDVLG 420
Db	378	IMDAVLITVCAIPGTAPON---KRNRSKGSIIYVFKPGHPEEVAFTNELFCRVEDLDD 434
Qy	421	LPQNTWKIGIMDEERTTVNLIKACIKAAARVVFINTGFLDRTGDDIHSTSMAGPMWRKG 480
Db	435	LPRLTKVGMDEERTSVNLDASIMEVADRLAFINTGFLDRTGDDIHSTSMAGANVRKA 494
Qy	481	TKWSQFWILAYEDHNVDAGLAAGSGRAQYKGMWMTTELMADVETKIAQPRAGASTAW 540
Db	495	DMCTAPWKQAYENNVVDAGIQRLPGRAQYKGMWMTTELMAEWKIKIQGPREGANTAW 554
Qy	541	VPSPTAATLHALHYHQVDVAAVOOGL-AGKERATIEQLLTIPAKELANAPDEIEEVDN 599
Db	555	VPSPTGATLHATHYHLVDVFKVQDELRAAGRDSLRNLTIPAPNTNNSEKKEEMDN 614
Qy	600	NQCSILGYVVRVWDQGVGCSKVPDIHDVVALMEDRATLRISQALLANLWRHGVTISADVRA 659
Db	615	NQCSILGYVVRVWEHGVGCSKVPDIHDIDLMEDRATLRISQMLANWIRHSDVVSQEVL 674
Qy	660	SUERWAPLDRONAGDVAYRNAPNFDSTAFLLAAQELILSGAQONGYTEPILHRRRE 719
Db	675	SUERWAVVVDKQAGDEAYRNAPNYDASLAFQAKDLIEGTYKSSGTYTEPILHRRRE 734
Qy	720	FRAR 723
Db	735	FRAK 738

CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbonyl compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX Sequence 712 AA;
 Query Match 59.1%; Score 2253; DB 22; Length 712;
 Best Local Similarity 61.2%; Pred. No. 3.5e-183;
 Matches 438; Conservative 89; Mismatches 183; Indels 6; Gaps 4;
 QY 9 LRIARVLYDFVNEALPGTIDPDPSFWAGVKKVADLTQONALLNARDELQAIQDKWHR 68
 DB 1 MQVAKVLYDFVTEAVLPRVGVDAEKFSWGAARADLTTPNRRELLAERDELQMLIDYHR 60
 QY 69 RRVIEPIMDAYRFLTEIGVLLPEDDFTITTSVDAEITTTAGPOLVVPVFNARFALN 128
 DB 61 NN-SGTDQEAEDFLKEIGVLEPEAAEIRTONVDEISSTAGPOLVVPVFNARFALN 119
 QY 129 AANARWGLSLYDALYGTDPVPTDCAEKGPTVKNVKGDKVAYARKFLDDSVLSSSGFGD 188
 DB 120 AANARWGLSLYDALYGTDPVPTDCAEKGPTVKNVKGDKVAYARKFLDDSVLSSSGFGD 179
 QY 189 ATGTVDQGVLPALPKSTGLANPGQFAGYTGAAESPTSVLLINHLHLEIILDPESQV 248
 DB 180 VEKYNITDGLAAHIGSVYLRKRESYRGFTGNFLDPEAILLETNGLHLEIILDPVHPI 239
 QY 249 GTTRPAGVKDVLISATITMDPDSVAAYDAADKVLGYRNWGLNGLKGLDAAVDDGTA 308
 DB 240 GKADKTKGLDVLISATITMDPDSVAAYDAADKVLGYRNWGLNGLKGLDAAVDDGTA 299
 QY 309 FLRVNLRDNTYAPGGGQFTLPGRSLMFRVNVGHMTNDAIVDTDGSSEVFEGIMDALFTG 368
 DB 300 FTRELKDRVYIGNGTELVLHGRSLLFVRNVGHLMQNPSIL-IDGEEIFEGIMDAVLT 358
 QY 369 LIAHGLKASDVNGPLNSRTGSIYIYKPKHGAFAVAFCELPFSRVEDVLGLPQNTMKI 428
 DB 359 VCAIPGIAFQV---KMRNSRKGSIYIYKPKHGAFAVAFCELPFSRVEDVLGLPQNTMKI 415
 QY 429 GIMDEERTTNNLKACIAAADRVVFTNTGFLDRTGDEIHTSMAGMVRKMGTKMSQPMI 488
 DB 416 GIMDEERTTNNLKACIAAADRVVFTNTGFLDRTGDEIHTSMAGMVRKMGTKMSQPMI 475
 QY 489 LAYEDHNVDAAGLAFSGRAQVKGMMTWMTBELMDVETKIAQPRAGASTAWPSPAT 548
 DB 476 QAYENNVNDAITQGLPQKAIQIGMGMWMTBELMDVETKIAQPRAGASTAWPSPAT 535
 QY 549 LHALHYHVDVAAVQOGL-ACKRRATIEQLITIPLAKELAWAPDEIREEDVNNCQSILGY 607
 DB 536 LHALHYHVDVAAVQOGL-ACKRRATIEQLITIPLAKELAWAPDEIREEDVNNCQSILGY 595
 QY 608 VYRVDOGVGSKVPDIHVALMEDRATFLRISQQLANWLPHGVITISADVRASLERVAPL 667
 DB 596 VYRVVEHGVGSKVPDIHVALMEDRATFLRISQQLANWLPHGVITISADVRASLERVAPL 655
 QY 668 VDRQNAVGVAPMPANFDDSIATFLAQELILSGAQQPNGVTEPILHRRRREPKAR 723
 DB 656 VDRQNAVGVAPMPANFDDSIATFLAQELILSGAQQPNGVTEPILHRRRREPKAR 711

RESULT 8
 ABG18357
 ID ABG18357 standard; Protein; 1190 AA.
 AC ABG18357;
 DT 18-FEB-2002 (first entry)
 XX Novel human diagnostic protein #18348.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 XX N-PSDB; AAS82544.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID No 48716; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (II) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIDO
 CC at ftp.wido.int/pub/published_pct_sequences.
 XX Sequence 1190 AA;
 Query Match 26.0%; Score 991.5; DB 22; Length 1190;
 Best Local Similarity 53.1%; Pred. No. 3.6e-75;
 Matches 195; Conservative 66; Mismatches 97; Indels 9; Gaps 3;
 QY 56 RDELOAQIDKWHRRVPIEDMDAYRQFLTEIGVLLPEDDFTITTSVDAEITTTAGPQ 115
 DB 187 RDRIQAALDEWHRRNPGPVKDKAAYKSFRLGELVLPQPERVTVETTGIDSEITSQAGPQ 246
 QY 116 LVVPLNARFALNANARWGLSLYDALYGTDPVPTDCAEKGPTVKNVKGDKVAYARKFL 175

Db 247 LVVPAMNARYALNAANRWGSLYDALYSDIIPQEGAMVSG--YDPQGEQVIAWRRFL 304
QY 176 DSVFLPSGSGFDATGFTVQDGLVVALPD-KSTGLANPGOPAGVTGAESPTSLLINH 234
Db 305 DESLPLENGSYDVVAFKVDVQQLRIQLKNGKEITLRTPAQFVGVGRGDAAPTCLILKNN 364
QY 235 GLHIEILIDPSQVQTTDRAGVKOVILSATTIMDFEDSVAADAAKVLGYRNWGLN 294
Db 365 GLHIGVQIDANGRIYKQDPFGHINDGIVETAISTILDCEDSVAADAEKILLERNLLGLM 424
QY 295 XGDIAAAVDKOGTAFRLVLRNDRNTAPGGGQFTLPGRSLMFRNVGHMTNDALVDTG 354
Db 435 QOTLQEKKEKNGRQIVRKLNDRHTAADGSEISLHGRSLFIRNVGHMTIPVWDSG 484
QY 355 SEVFEGINDALFTGHAHGLKASDVNGPLINSRTGSIYIVKPKKHGPAEVAFTCELSFR 414
Db 485 NEIPEGILDVGMTGAIALYDLKVQK-----NSRTGSIYIVKPKKHGQOEVAFAFANKLFR 538
QY 415 VEDVLGL 421
Db 539 IETMLGI 545

RESULT 9
ABG25453
ID ABG25453 standard; Protein; 830 AA.
XX AC ABG25453;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #25444.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS89640.
XX CC New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX PS Claim 20; SEQ ID No 55812; 103pp; English.

The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 830 AA;
Query Match 19.5%; Score 741.5; DB 22; Length 830;
Best Local Similarity 60.3%; Pred. No. 4.5e-54;
Matches 146; Conservative 34; Mismatches 53; Indels 9; Gaps 3;
QY 412 FSRV--EDVLGLPQNTWKIGINDEERTTVNLKACIKAAADRVVFTNGFLDRTGDEIHT 469
Db 398 FSKMAGHDTIVLGTETVYKMGINDEERTSLNLESCIAQARNRVAFINTGFLDRTGDEHHS 457
QY 470 SMEAGPMVRKGTMKSQPWILAYEDHNVNDAAGAFSGRAQVKGVMWMTMELMADVYETKI 529
Db 458 VMEAGPMVRKQNMKSTPWIKAYERNVLSGLFCGLRGKQIKGKMWAMPDLADYVSKG 517
QY 530 AQRAGASTAWVSPPTAATLHALHYHOVDVAVQOGLA-----GKRRTATIEQLLTPLAK 584
Db 518 DQLRAGANTAWVSPPTAATLHALHYHQTNNVQSVQANIAQTEFNABEPEPLIDDLTTPVAE 577
QY 585 ELAWAPEIREEVDNNQCSTILGYVWRWVQGVCSKVPDIHDVVALMEDRATLRISSQLA 644
Db 578 NANWSAQETQOELDNNVQGITLVYVWRVQAGTCCKVPDIHNVVALMAARATLSSISH--A 635
QY 645 NW 646
Db 636 EW 637

RESULT 10
ABG18358
ID ABG18358 standard; Protein; 504 AA.
XX AC ABG18358;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #18349.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS82545.
XX CC New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

```

XX PS Claim 20; SEQ ID No 48717; 103pp; English.
XX PS
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 504 AA;
XX
XX Query Match 15.6%; Score 594.5; DB 22; Length 504;
XX Best Local Similarity 50.9%; Pred. No. 7.6e-42;
XX Matches 115; Conservative 40; Mismatches 68; Indels 3; Gaps 2;
XX
XX QY 9 LRIRARLVDFVNNALPCTDIDPSFWAGVDKVVADLTQNCALLNARDLQADKWHR 68
XX DB 139 LRIDANFRFVDEVLPGTGLDAAAFWRNFEIYHDLAPENRQLLAERDRIQALDSEHR 198
XX
XX QY 69 RRVIEPIDMDAYROFLTEIGYLLPEPDFTITTSVDAAEITTTAGPQLVVPVNLARFALN 128
XX DB 199 SNPGFVKDKAAYKSFLELGYLVFPQPERVITETGIDSEITSQAGPQLVVPANARYALN 258
XX
XX QY 129 AANARWGLYDALYGTVDVITPDGAEGKPTYNKVRGDKVIAYARKFLDDSVPLSSGSPGD 188
XX DB 259 AANARWGLYDALYGSIIIPQEGAMVSG--YDFQGEQVIAWRRFLDESPLPENGSYQD 316
XX
XX QY 189 ATGFTVODGQLVWALPD-KSTGLANPGQFAGYTGAAESPTSVLLIN 233
XX DB 317 VVAFKVLQ-----IDANGRI 362
XX
XX RESULT 11
XX ABG18356
XX ID ABG18356 standard; Protein; 1641 AA.
XX XX
XX AC ABG18356;
XX XX
XX DT 18-FEB-2002 (first entry)
XX XX
XX DE Novel human diagnostic protein #18347.
XX XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX XX
XX PN WO200175067-A2.
XX XX
XX PD 11-OCT-2001.
XX XX
XX XX 30-MAR-2001; 2001WO-US08631.
XX EF
XX XX 31-MAR-2000; 2000US-0540217.
XX PR
XX FR 23-AUG-2000; 2000US-0649167.
XX XX
XX PA (HYSE-) HYSEQ INC.

```

```

XX PI Drmanac RT, Liu C, Tang YT;
XX XX
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS82543.
XX XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX XX
XX PS Claim 20; SEQ ID No 48715; 103pp; English.
XX XX
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1641 AA;
XX
XX Query Match 15.4%; Score 585.5; DB 22; Length 1641;
XX Best Local Similarity 42.4%; Pred. No. 2.6e-40;
XX Matches 122; Conservative 46; Mismatches 71; Indels 49; Gaps 3;
XX
XX QY 9 LRIRARLVDFVNNALPCTDIDPSFWAGVDKVVADLTQNCALLNARDLQADKWHR 68
XX DB 942 LRIDANFRFVDEVLPGTGLDAAAFWRNFEIYHDLAPENRQLLAERDRIQALDSEHR 1001
XX
XX QY 69 RRVIEPIDMDAYROFLTEIGYLLPEPDFTITTSVDAAEITTTAGPQLVVPVNLARFALN 128
XX DB 1002 SNPGFVKDKAAYKSFLELGYLVFPQPERVITETGIDSEITSQAGPQLVVPANARYALN 1061
XX
XX QY 129 AANARWGLYDALYGTVDVITPDGAEGKPTYNKVRGDKVIAYARKFLDDSVPLSSGSPGD 188
XX DB 1062 AANARWGLYDALYGSIIIPQEGAMVSG--YDFQGEQVIAWRRFLDESPLPENGSYQD 1119
XX
XX QY 189 ATGFTVODGQLVWALPD-KSTGLANPGQFAGYTGAAESPTSVLLINHGHLHILIDPSQV 248
XX DB 1120 VVAFKVLQ-----IDANGRI 1133
XX
XX QY 249 GTTDRAGVKDVILSAITTIMDFDSVAADKVLGYRNWLGKNG 296
XX DB 1134 GKNDPAHINDVIVEAAISTILDENSGRAVNPFIKSV-VPHLLGLING 1180
XX
XX RESULT 12
XX ABG18355
XX ID ABG18355 standard; Protein; 441 AA.
XX XX
XX AC ABG18355;
XX XX
XX DT 18-FEB-2002 (first entry)
XX XX
XX DE Novel human diagnostic protein #18346.
XX XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

```

KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 XX N-PSDB; AAS82542.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID No 48714; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG030377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 441 AA;
 SQ
 Query Match 11.4%; Score 433.5; DB 22; Length 441;
 Best Local Similarity 55.9%; Pred. No. 3.4e-28;
 Matches 81; Conservative 26; Mismatches 37; Indels 1; Gaps 1;
 QY 9 LRIARVLYDFVNEALPGTDIDPDSFWAGVKVADLTTPQNALNARDELQAKDKWHR 68
 DB 283 LRIDANKEFVDEVLPGTGLDAARFNRFDIVHDLAPENQLLAERDQIAALDEWHR 342
 QY 69 RRVEIPDMNARQFTEITGYLLPEPDDTTITSGVDABITTAGPOLVVPVNLNARFALN 128
 DB 343 SNFGPVKDKAAYKSFRLGVLVPOPERVTVTTGIDTSITSGAGPOLVVPANARYALN 402
 QY 129 AANARWGSLYDALYGDVVPETDGA 153
 DB 403 AANARWGSLYDALYGSIIIPQ-EGA 426
 RESULT 13
 ABP03007
 ID ABP03007 standard; Protein; 130 AA.
 XX
 AC ABP03007;
 XX 24-JUN-2002 (first entry)
 XX Human ORFX protein sequence SEQ ID NO:5996.
 XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.
 XX Homo sapiens.
 XX WO200192523-A2.
 XX 06-DEC-2001.
 XX 29-MAY-2001; 2001WO-US10836.
 XX 30-MAY-2000; 2000US-206132P.
 XX 29-AUG-2000; 2000US-228716P.
 XX (CURA-) CURAGEN CORP.
 XX Shimkets RA, Leach MD;
 XX WPI: 2002-106308/14.
 XX N-PSDB; ABN18759.
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders -
 XX Disclosure; SEQ ID 5996; 1037pp; English.
 XX The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis.
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 130 AA;
 SQ
 Query Match 11.0%; Score 420; DB 23; Length 130;
 Best Local Similarity 59.7%; Pred. No. 7.9e-28;
 Matches 77; Conservative 24; Mismatches 28; Indels 0; Gaps 0;
 QY 1 TDRVSVGNLRATRVLYDFVNEALPGTDIDPDSFWAGVKVADLTTPQNALNARDELQ 60
 DB 2 TERVQGGSLQVAKVLYDFVVEEAVFGAGVDAETFWAGDKVITELAPNRLAKRDLQ 61

QY 61 AQIDKWHRRVIEBIDMAYRQFLTEIGYLLPSPDDFTITTSVGDABEITTTAGPQLWVPV 120
 Db 62 AQIDQWHRDHRASLEAVAKFLQEGYLLPTPAEPSVNTANVDTEITTAGPQLWVPI 121

QY 121 LNARFALNA 129
 Db 122 LNARFALNA 130

RESULT 14
 AAR37585
 ID AAR37585 standard; Protein; 533 AA.
 XX AAR37585;
 AC AAR37585;
 XX 25-MAR-2003 (updated)
 DT 29-JUL-1993 (first entry)
 XX E.coli malate synthase expressed from MTAceB2 construct.
 XX Growth Hormone; metallothionein Ia; inducible promoter;
 KW bacterial gene expression; malate synthase; glyoxylate cycle;
 KW transgenic animal.
 XX Escherichia coli.
 OS WO9218635-A1.
 XX 29-OCT-1992.
 PD 13-APR-1992; 92WO-AU00164.
 XX 16-APR-1991; 91AU-0005664.
 PR (CSIR) COMMONWEALTH SCI & IND RES ORG.
 XX Brownlee AG, Nancarrow CD, Ward KA;
 FI WPI; 1992-382118/46.
 DR N-PSDB; AAQ43182.
 XX Gene expression cassette for high level expression of bacterial
 PT and/or plant genes - comprises inducible promoter and
 PT 3'-non-coding sequence of growth hormone exon 5
 XX Example 2; Fig 6; 8pp; English.
 PS Construct MTAceB2 consists of the sheep metallothionein-Ia gene
 CC promoter sequence joined to the coding sequence of the E.coli aceB
 CC gene at a unique SalI site. This sequence was then joined by the
 CC 3' sequence of exon 5 of sheep growth hormone gene. Minor sequence
 CC modification around the initiation codon of the bacterial aceB gene
 CC was made by site-directed mutagenesis. The construct was transferred
 CC to transgenic mice and the mice were analysed for expression of
 CC malate synthase. The value for expression (nmoles product formed/mg
 CC protein/20 min) was 34.3, c.f. 0 in the control cell line.
 CC See AAQ43177-Q43183.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX Sequence 533 AA;

Query Match 5.0%; Score 189.5; DB 13; Length 533;
 Best Local Similarity 22.7%; Pred.No. 3.le-07;
 Matches 120; Conservative 67; Mismatches 253; Indels 89; Gaps 21;

QY 144 TDVIPETDGAEGKPTYNKRGDKVIAYARKFLDSDVPLSSGFGDATGFTVQDGLWVAL 203
 Db 2 TEQATTTDELAFTEPYGEQEQKILTAFAVEFLTELVE-----THPTQENKLLAAR 51
 QY 204 PDKSTGLANPQFAGYTGAAESPTSVLLINHLHIEILDPSQVGTTRAGVKVILE 263
 Db 52 IQQQQDIDN-GTLPDFDIFSETASIRDADWKIRGIPADLEDRRVEITGTVKRWVINALNAN 110

QY 264 AITTIMPDEDSVAADAAADKVLGYRNMLGLNKGDLAAAVDKDGTAFLE-VLNDRDNYTAP 322
 Db 111 VKVFADPEDSLAP-----DW---NK-----VIDGQINLRDAVNGTISYTN 149

QY 323 GGGQFTLPGRSUFVRNVGHLMTNDIAVDTDGSEVFEGIMDALET-GLIAIHGKA--SD 379
 Db 150 AGKIYQLKPNPAVLICVRGJHLPEKHVTWEG-----EAIPGSLDFDLYFFHNYQALLAK 205

QY 380 VNGPLINSRTGSIYIVKPKMHGPAEVAFTCELFGRVEDVLGLPQNTMKIGIMDEERTTV 439
 Db 206 GSGP-----YFYLPTQSQWQEAAMWSEVFSVAEDRENLPRTIKATLLIETLPAVF 256

QY 440 NLKACIKAAADRVVFINTGFLDRTGDEIHTSMEAGPMV---RKGTMSQPMILAYED--- 493
 Db 257 QMDEILHALRDHIIVGLNCGRWDIYFSYIKTLKPNVDRVLPDRQAVTWDKPFLNAYSRLI 316

QY 494 ---HNVDA---GLAAGPSGRAQVKGWMTMTLMADMVETKIAQPRAGASTAWPSPSTA 546
 Db 317 KTKHKGAFAMGKMAAFIPSKDEHNQ-VLNKVKAD-----KSLEANNHGDGTWIAHGL 371

QY 547 ATLHALHYHQVDA-AVQOGLAGKRRATIEOL--LTIPLAKELAWP---DEIRFEVDNN 600
 Db 372 A-----DTAMAVENDILGSRKNQLEYMREQDAPITADQLLAPCDGERTEGMRAN 421

QY 601 COSILGYVVRVWDQGVGCKVPDIHDVALMEDRATLRISSQLLANWLRLH 649
 Db 422 IRVAVQYIEAMI-SONGCVPI-----YGLMEDATAEISRTSIQWNIH 464

RESULT 15
 AAO16702
 ID AAO16702 standard; Protein; 533 AA.
 XX AAO16702;
 AC AAO16702;
 XX 29-MAY-2003 (first entry)
 DT Escherichia coli aceB protein.
 DE Enzyme; L-amino acid preparation; malate synthase A; aceB; L-threonine;
 KW human medicine; pharmaceutical industry; foodstuffs; animal nutrition.
 XX Escherichia coli.
 OS WO2003008604-A2.
 XX 30-JAN-2003.
 PD 03-JUL-2002; 2002WO-EP07352.
 XX 18-JUL-2001; 2001DE-1035051.
 PR 23-JUL-2001; 2001US-306867P.
 XX (DEGS) DESUSSA AG.
 XX Hermann T;
 PI WPI; 2003-248014/24.
 DR N-PSDB; AAL52112.
 XX Preparing L-amino acids, e.g. L-threonine, by fermenting microorganisms
 PT of Enterobacteriaceae family in which the aceB gene is attenuated in
 PT particular eliminated, and isolating L-amino acid from culture medium -
 XX Disclosure; Page 31-33; 36pp; English.
 PS The invention comprises a method of preparing L-amino acids. The method
 CC involves fermenting microorganisms of Enterobacteriaceae family that
 CC produce the desired L-amino acid, in which the aceB (malate synthase A)
 CC gene is eliminated. The method is useful for preparing L-amino acids
 CC (e.g. L-threonine). The L-amino acids produced by the method of the
 CC invention are useful in human medicine and in the pharmaceuticals

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 28.8617 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-2

Perfect score: 3810

Sequence: 1 TDRVSGNLRARVLYDFVN.....KARAERKPSDRAGDDAAR 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PTC_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PTCUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	100.0	748	11 US-09-272-975-2	Sequence 2, Appli
2	3810	100.0	753	11 US-09-272-975-58	Sequence 58, Appl
3	2289	60.1	739	10 US-09-738-626-6988	Sequence 6988, Ap
4	184	4.8	530	15 US-10-158-761-9582	Sequence 9582, Ap
5	145.5	3.8	541	15 US-10-158-761-9538	Sequence 9538, Ap
6	135	3.5	3816	11 US-09-808-880-3	Sequence 3, Appli
7	132.5	3.5	5245	12 US-10-329-079-45	Sequence 45, Appl
8	130	3.4	1624	16 US-10-080-170-539	Sequence 539, App
9	128	3.4	878	9 US-09-905-983-48	Sequence 48, Appl
10	128	3.4	878	14 US-10-165-049-2	Sequence 2, Appli
11	128	3.4	878	14 US-10-165-049-3	Sequence 3, Appli
12	128	3.4	11877	10 US-09-861-289-6	Sequence 6, Appli
13	128	3.4	11877	10 US-09-860-846-6	Sequence 6, Appli
14	128	3.4	11877	11 US-09-836-821-6	Sequence 6, Appli
15	128	3.4	11877	12 US-10-271-889-49	Sequence 49, Appl

16	128	3.4	12199	11 US-09-988-384B-6	Sequence 6, Appli
17	127.5	3.3	5701	9 US-09-864-761-37319	Sequence 37319, A
18	127.5	3.3	5701	12 US-10-029-386-32438	Sequence 32438, A
19	127.5	3.3	26926	10 US-09-759-508B-2	Sequence 2, Appli
20	126.5	3.3	3739	10 US-09-861-289-33	Sequence 33, Appl
21	126.5	3.3	3739	10 US-09-860-846-33	Sequence 33, Appl
22	126.5	3.3	3739	11 US-09-988-384B-33	Sequence 33, Appl
23	126.5	3.3	3739	11 US-09-836-821-33	Sequence 33, Appl
24	126.5	3.3	3739	11 US-09-793-708-2	Sequence 2, Appli
25	126.5	3.3	3739	12 US-10-201-365-3	Sequence 3, Appli
26	126.5	3.3	3739	12 US-10-160-539-2	Sequence 33, Appl
27	126.5	3.3	3739	12 US-10-271-889-33	Sequence 7, Appli
28	125.5	3.3	2039	15 US-10-193-584-7	Sequence 7, Appli
29	124.5	3.3	3939	15 US-10-156-761-10434	Sequence 10434, A
30	124	3.3	1346	10 US-09-861-289-37	Sequence 37, Appl
31	124	3.3	1346	10 US-09-860-846-37	Sequence 37, Appl
32	124	3.3	1346	11 US-09-988-384B-37	Sequence 37, Appl
33	124	3.3	1346	11 US-09-836-821-37	Sequence 37, Appl
34	124	3.3	1346	12 US-10-271-889-37	Sequence 37, Appl
35	124	3.3	4999	10 US-09-976-059-14	Sequence 14, Appl
36	123	3.2	1346	11 US-09-793-708-4	Sequence 4, Appli
37	123	3.2	1346	12 US-10-201-365-5	Sequence 5, Appli
38	123	3.2	1346	12 US-10-160-539-4	Sequence 4, Appli
39	122	3.2	821	12 US-10-174-677-28	Sequence 28, Appl
40	122	3.2	4999	10 US-09-976-059-15	Sequence 15, Appl
41	119.5	3.1	7257	14 US-10-014-717-5	Sequence 5, Appli
42	119	3.1	642	15 US-10-156-761-9781	Sequence 9781, Ap
43	118.5	3.1	915	12 US-10-193-764-35	Sequence 35, Appl
44	118.5	3.1	1222	12 US-10-193-764-37	Sequence 37, Appl
45	118.5	3.1	1228	12 US-10-193-764-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-09-272-975-2
; Sequence 2, Application US/09272975
; Publication No. US2003002774A1
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Ronald C.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: TUBERCULOSIS ANTIGENS AND METHODS
; FILE REFERENCE: 210121.474
; CURRENT APPLICATION NUMBER: US/09/272,975
; CURRENT FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 748
; ORGANISM: Mycobacterium tuberculosis
US-09-272-975-2

Query Match 100.0%; Score 3810; DB 11; Length 748;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDRVSGNLRARVLYDFVNNEALPGTDIDPSWAGVKVADLTQONQALLNARDELQ 60
Db 9 TDRVSGNLRARVLYDFVNNEALPGTDIDPSWAGVKVADLTQONQALLNARDELQ 68
QY 61 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDTTITTSQVDAEITTTAGPOLVVPV 120
Db 69 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDTTITTSQVDAEITTTAGPOLVVPV 128
QY 121 LNARPALNAARVGSYDLYGTDVIPETDGAEGKPTYNKVRGDKVIYARKFLDDSV 180
Db 129 LNARPALNAARVGSYDLYGTDVIPETDGAEGKPTYNKVRGDKVIYARKFLDDSV 188
QY 181 LSSSGFGDGTGTVDQGVVVALPKSTGLANPGQFAGYTGAEPSPTSVLLINHLHLEI 240

```
Db 189 LSSGSGDATGFTVQDGLVVALPKSTGLANPQFAGYTGAAESPTSVLLINHLHIEI 248
Qy 241 LIDPESQVGTTRAGVKDVILLESATITIMDPEDSVAADKVLGYRNWLGKNGDLAA 300
Db 249 LIDPESQVGTTRAGVKDVILLESATITIMDPEDSVAADKVLGYRNWLGKNGDLAA 308
Qy 301 AVDXDGTAFRLVLRNDRNTAFPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDSEVFE 360
Db 309 AVDXDGTAFRLVLRNDRNTAFPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDSEVFE 368
Qy 361 IMDALFTGLIAIHGKASDVNGPLINSRTGSIYIVKPMHGPAEVAFTCELFSEVFE 420
Db 369 IMDALFTGLIAIHGKASDVNGPLINSRTGSIYIVKPMHGPAEVAFTCELFSEVFE 428
Qy 421 LPQNTMKIGINDEERRITVNLKACIKAAADRVFINTGFLDRTGDEIHTSMEAGPMVRK 480
Db 429 LPQNTMKIGINDEERRITVNLKACIKAAADRVFINTGFLDRTGDEIHTSMEAGPMVRK 488
Qy 481 TMKSQPWILAYEDHNVDAGLAAGSFGRAQVKGKGMWMTMELMADMVETKIAQPRAGASTAW 540
Db 489 TMKSQPWILAYEDHNVDAGLAAGSFGRAQVKGKGMWMTMELMADMVETKIAQPRAGASTAW 548
Qy 541 VPSPTAATLHALHGHQVDVAAVQOGLAGKRRATIEQLTTIPLAKELAWAPDEIREVDNN 600
Db 549 VPSPTAATLHALHGHQVDVAAVQOGLAGKRRATIEQLTTIPLAKELAWAPDEIREVDNN 608
Qy 601 CQSILGYVVRVWDQGVGCKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 660
Db 609 CQSILGYVVRVWDQGVGCKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 668
Qy 661 LERMAPLVDRONAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQPNGYTEPILHRRRREF 720
Db 669 LERMAPLVDRONAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQPNGYTEPILHRRRREF 728
Qy 721 KARAAEKPAFSDRAGDDAAR 740
Db 729 KARAAEKPAFSDRAGDDAAR 748

RESULT 2
US-09-272-975-58
; Sequence 58, Application US/09272975
; Publication No. US2003002774A1
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Ronald C.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: TUBERCULOSIS ANTIGENS AND METHODS
; FILE REFERENCE: 21021 474
; CURRENT APPLICATION NUMBER: US/09/272,975
; CURRENT FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-272-975-58

Query Match 100.0%; Score 3810; DB 11; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDRSVGNLRIARVLYDFVNEALPGTDIDPDSFAGVGVKVAADLPQNALNARDELQ 60
Db 14 TDRSVGNLRIARVLYDFVNEALPGTDIDPDSFAGVGVKVAADLPQNALNARDELQ 73
Qy 61 AQIDKWHRRVIEPIDMDAYRQELTEIGYLLPPDPDFTITTSVGDAEITTTAGPQLVWVP 120
Db 74 AQIDKWHRRVIEPIDMDAYRQELTEIGYLLPPDPDFTITTSVGDAEITTTAGPQLVWVP 133
```

```
Qy 121 LNARFALNANARWGLSDALYGTQDVIPEITDCAEKGPTYNKVRGDKVIAYARKFLDSDVP 180
Db 134 LNARFALNANARWGLSDALYGTQDVIPEITDCAEKGPTYNKVRGDKVIAYARKFLDSDVP 193
Qy 181 LSSGSGDATGFTVQDGLVVALPKSTGLANPQFAGYTGAAESPTSVLLINHLHIEI 240
Db 194 LSSGSGDATGFTVQDGLVVALPKSTGLANPQFAGYTGAAESPTSVLLINHLHIEI 253
Qy 241 LIDPESQVGTTRAGVKDVILLESATITIMDPEDSVAADKVLGYRNWLGKNGDLAA 300
Db 254 LIDPESQVGTTRAGVKDVILLESATITIMDPEDSVAADKVLGYRNWLGKNGDLAA 313
Qy 301 AVDXDGTAFRLVLRNDRNTAFPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDSEVFE 360
Db 314 AVDXDGTAFRLVLRNDRNTAFPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDSEVFE 373
Qy 361 IMDALFTGLIAIHGKASDVNGPLINSRTGSIYIVKPMHGPAEVAFTCELFSEVFE 420
Db 374 IMDALFTGLIAIHGKASDVNGPLINSRTGSIYIVKPMHGPAEVAFTCELFSEVFE 433
Qy 421 LPQNTMKIGINDEERRITVNLKACIKAAADRVFINTGFLDRTGDEIHTSMEAGPMVRK 480
Db 434 LPQNTMKIGINDEERRITVNLKACIKAAADRVFINTGFLDRTGDEIHTSMEAGPMVRK 493
Qy 481 TMKSQPWILAYEDHNVDAGLAAGSFGRAQVKGKGMWMTMELMADMVETKIAQPRAGASTAW 540
Db 494 TMKSQPWILAYEDHNVDAGLAAGSFGRAQVKGKGMWMTMELMADMVETKIAQPRAGASTAW 553
Qy 541 VPSPTAATLHALHGHQVDVAAVQOGLAGKRRATIEQLTTIPLAKELAWAPDEIREVDNN 600
Db 554 VPSPTAATLHALHGHQVDVAAVQOGLAGKRRATIEQLTTIPLAKELAWAPDEIREVDNN 613
Qy 601 CQSILGYVVRVWDQGVGCKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 660
Db 614 CQSILGYVVRVWDQGVGCKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 673
Qy 661 LERMAPLVDRONAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQPNGYTEPILHRRRREF 720
Db 674 LERMAPLVDRONAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQPNGYTEPILHRRRREF 733
Qy 721 KARAAEKPAFSDRAGDDAAR 740
Db 734 KARAAEKPAFSDRAGDDAAR 753

RESULT 3
US-09-738-626-6988
; Sequence 6988, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6988
```



```
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6988

Query Match      60.1%; Score 2289; DB 10; Length 739;
Best Local Similarity 61.3%; Pred. No. 4.3e-205;
Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;

QY 1  TDRVSGNLRARLVYDFYNNELPGTDIDPDSFVAGVQKVADLTPOQALLNARDELQ 60
DB 20  TERVDAGGQVAKLVYDFTEAVLFRVGDVDAKFWGFAAIARDLTPRRELLARDELQ 79

QY 61  AQIDKWHRRVPIEDMDAYROFLTEIGYLLPEPDDFTITTSVDAEITTTAGPOLVVPV 120
DB 80  MLIDDYHRNN-SGTDIQEAYEDFLKEIGYLVPEPEAAEIRIQNVDEITSSTAGPOLVPEI 138

QY 121  LNARFALNANRWGSLYDALYGTVDIPETDGAEGKPTYNKVRGKQVAYARKFDDSDVP 180
DB 139  LNARFALNANRWGSLYDALYGTVNAIPETDGAEGKPTYNKVRGKQVAYARKFDDSDVP 198

QY 181  LSSGSGDATGFTVQDQGLVVALPKSTGLANPGOPAGVTGAESPTSLLINHLGHIEI 240
DB 199  LDGASHADVKEYNITDGLAAHIGDSVYELKXNRESYRGFTGNFLDPEAILLETNGHLIEL 258

QY 241  LIDPESQVGTTRAGVKDVLISAITTIMDFDSVAADVADAKVLGYRNWLGKNGDLAA 300
DB 259  QIDPVHPIGKADKTLGKDIVLESAITTIMDFDSVAADVADAKVLGYRNWLGKNGDLAA 318

QY 301  AVDKDGTAFRLVRNDRNTVAPGGQFTLPGRSLMFRVNVGHMTNDAIVDTDGSEVPEG 360
DB 319  EMSKNGRIETRELNKDRVIGRNGTELVLHGRSLIFVRNVGHLMQNPSIL-IDGEEIPEG 377

QY 361  IMDALTGLIAHGLKASDVNGPLINSRTGSYIVKPKMHGPAEVAFTCELSRVEDVLG 420
DB 378  INDVAVLTTCATPGIAPQN---KMRNSRKGSYIVKPKMHGPAEVAFTCELSRVEDVLG 434

QY 421  LPQNTMKIGIMDEERTTNVLRACIAAADRVVFNITGFLDRTGDEIHTSMGAPVWRKG 480
DB 435  LPRHTLVGVNDERTSVNLDASINEVADRLAFINTGFLDRTGDEIHTSMGAPVWRKA 494

QY 481  TKKSQPWILAYEDHNVDAGLAAGFSGRAQVKGGMVTMTLMDVMTETKIAQPRAGASTAW 540
DB 495  DMQTAFWKQAYENNVDAGIQRLPGKAOIGKGMWMTLMDVMTETKIAQPRAGASTAW 554

QY 541  VPSPTAATHLHYHOVDVAQVQGL-AGKRRATIEQLITIPLAKELAWAPDEIREVDN 599
DB 555  VPSPTGATLHATHYLVVDVFKVQDELRAAGRDSLRNLTITPTAPNTNWSSEKKEEMDN 614

QY 600  NCOSILGYVVRVVDQGVGSKVPDIHDVALMEDRATLRISQALLANLWLRHGVITGADVRA 659
DB 615  NCOSILGYVVRVVEHGVGSKVPDIHDIDLMEDRATLRISQALLANLWLRHGVITGADVRA 674

QY 660  SLERMAPLVDRQVAGVAVRMAPNFDSDIAFLAAQELILSGAQQPNQNGVTEPILHRRRE 719
DB 675  SLERMAPLVDRQVAGVAVRMAPNFDSDIAFLAAQELILSGAQQPNQNGVTEPILHRRRE 734

QY 720  FKAR 723
DB 735  FKAR 738
```

RESULT 4

```
US-10-156-761-9582
; Sequence 9582, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
```

RESULT 5

```
US-10-156-761-9538
; Sequence 9538, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
```

```
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9582
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9582

Query Match      4.8%; Score 184; DB 15; Length 530;
Best Local Similarity 21.9%; Pred. No. 4e-08;
Matches 124; Conservative 74; Mismatches 245; Indels 124; Gaps 23;

QY 197  GOLVVALLPKSTGLANPGOPAGVTGAESPTSLLINHLGHIEI-----LIDPES 246
DB 33  GRLVAAPGERPELLKERRRQAIRLAAGSALDPELATAGVRADASWRVAAPAPGLTDRV 92

QY 247  QV-CTTRAGVKDVLISAITTIMDFDSVAADVADAKVLGYRNWLGKNGDLAAAVDKD 305
DB 93  BITPPORRTAVNALSKARVMADFDATSP-----TWNIIQGQL----- 134

QY 306  GTAFRLVRNDRNTVAPGGQFTLPGR-SLMFVRNVGHMTNDAIVDTDGSEVEGIMDA 364
DB 135  --TLDAIERRIDFTTPEGKEYRLGDSLATIWRPCWHLTEHL-EFDGRPVASLVD- 190

QY 365  LFTGLIAIH-GLKASDV-NGPLINSRTGSYIVKPKMHGPAEVAFTCELSRVEDVLGP 422
DB 191  --FGLYFFHCAQRQIDAGHGP-----YFVLPKLENRYEARLWVDFVLAQELIGIP 239

QY 423  QNTWKIGIMDEERTTNVLRACIAAADRVVFNITGFLDRTGDEIHTSMGAPVWR--RK 479
DB 240  PGTRATVLTITTAAVEBEILYELRHSAGLNAGWDVILSLIKTFGHRDTFLPDRA 299

QY 480  GTMSQPWILAYED-----HNVDAGLAAGFSGRAQV-KGGMVTMTLMDVMTETKIAQ 532
DB 300  TVMTAPFLRAYTELLVTRCHRRGAHAIGGMS--AQVFGRPDPAAGAAALAKVRLDKREA 357

QY 533  RAGASTAVPSPTAATHLHYHOVDVAQVQGLAGKERATIEQ-----LLTIP 581
DB 358  ECGFDGGSWAHPGLVPM-----CREVFDGVIGRPHQIETRDDVEVTAADLLSV- 407

QY 582  LAKELAWAPDEIREVDNCCOSILGYVVRVVDQGVGSKVPDIHDVALMEDRATLRISQ 641
DB 408  --RRISGPPSP--EGVATNIAVSLRYFDAML-RGRGAVAL-----YGLMEDAATAETARV 457

QY 642  LLANLWLRHGVITSDVADVASLERMAPLVDRQVAGVAVRMAPNFDSDIAFLAAQELILSG 701
DB 458  QIOWMLRHRV-----VDR-----DTVVRLLDDELAALG 485

QY 702  AQQPNQNGVTEPILHRRREFKARAAEKP 728
DB 486  AEYFWAALDEV--RALFERTALAAELP 510
```

;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-262
;; CURRENT APPLICATION NUMBER: US/10/156,761
;; CURRENT FILING DATE: 2002-05-29
;; PRIOR APPLICATION NUMBER: JP 2001-204089
;; PRIOR FILING DATE: 2001-05-30
;; PRIOR APPLICATION NUMBER: JP 2001-272697
;; PRIOR FILING DATE: 2001-08-02
;; NUMBER OF SEQ ID NOS: 15109
;; SEQ ID NO 9538
;; LENGTH: 541
;; TYPE: PRT
;; ORGANISM: Streptomyces avermitilis
US-10-156-761-9538

Query Match 3.8%; Score 145.5; DB 15; Length 541;
Best Local Similarity 20.1%; Pred. No. 0.00017;
Matches 101; Conservative 67; Mismatches 210; Indels 125; Gaps 21;

QY 249 GTTDRAGKDVILSAITTIMDFDSVAANDADKV.GYRNWGLNKGDLAAAVDKGTA 308
Db 96 GPTDKMTINALSGARWLADFEDAGAP-----TWENVVYQGL-----N 135
QY 309 FLRVLRNDRNTAGGGG-FTLPGR--SLMFVENVGLMTINDAIVTDGSEVFEGIMDA 364
Db 136 LTDAYTRNIDFTDPKSGSYALKANEELATVVMRPGWHLNRHLVDSQGRPVFGALVD- 194
QY 365 LFTGLIAHGLKASDVNGPLNSGTGIYIVKPKMHGPAEVAFTCELSRVEDVLGLPQN 424
Db 195 -FGLYFFHNARR-----LLDLGKGP-YFYLPTKTESHLEARLWNVFVPAQYVVGIPQ 245
QY 425 TMKIGIMDEERTTVNLKACIAKAADRVFINTGFLDRTGDEIHTSNEAGPMV-----RK 480
Db 246 TVRATVLETTTAAAYEMEEIYELRHDHAGLNAGRWDLFSIVKNFRDGGAKFVLPRNA 305
QY 481 TMKQPMILAYED-----HNVDAGLAGFSGRAQVQKGMWMTMTLMADMVETKI 529
Db 306 VTMTAPFMRYATELLVTRCHKRGAAH-GGMAAFIPSRDEEVKNVAFKVKAD-----KD 360
QY 530 AQRAGASTAWPPTAATLHALHYQVDVAAVQOGLAGKRATIEQLLIPLAKELAWA 589
Db 361 REAGDGFDSWVAHPDLV-----PIAMASFDVILGDK-----PNQK----- 396
QY 590 PDEIREVD-----NNQSTILGVVVRWVDGVGCSKVVDIHDV 627
Db 397 -DLREDVSVAPGLLIAIDSLAEKPTYNGLVNAVQVIRVIEAWL-RGLGAVAI-----F 449
QY 628 ALMEDRATLRISQOLLANWRHGVITSAVRA-----SLRMAPLVDRQAGDVAYR 679
Db 450 NLMEATAETASRSQIQWINAGVFEFENGSKATPELARKVAABELANL--RAELGGEAF- 506
QY 680 PMAPNFDDSI AFLAAQELILSGA 702
Db 507 -AAGWQQ-----AHDLLLVKA 522

RESULT 6

US-09-808-880-3
;; Sequence 3, Application US/09808880
;; Publication No. US20030027287A1
;; GENERAL INFORMATION:
;; APPLICANT: Betlach, Mary C.
;; APPLICANT: Shah, Sanjay Krishnakant
;; APPLICANT: McDaniel, Robert
;; APPLICANT: Tang, Li
;; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE

;; FILE REFERENCE: 30062-20029.00
;; CURRENT APPLICATION NUMBER: US/09/808,880
;; CURRENT FILING DATE: 2001-03-14
;; PRIOR APPLICATION NUMBER: US/09/428,517
;; PRIOR FILING DATE: 1999-10-28
;; PRIOR APPLICATION NUMBER: 60/120,254
;; PRIOR FILING DATE: 1999-02-16

;; PRIOR APPLICATION NUMBER: 60/106,100
;; PRIOR FILING DATE: 1998-10-29
;; NUMBER OF SEQ ID NOS: 12
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 3
;; LENGTH: 3816
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;; OTHER INFORMATION: Oleandolide PKS
US-09-808-880-3

Query Match 3.5%; Score 135; DB 11; Length 3816;
Best Local Similarity 21.3%; Pred. No. 0.04;
Matches 190; Conservative 112; Mismatches 297; Indels 294; Gaps 49;

QY 25 PG-----TDIDPD-SFWAGVDKVVADLTIPQALLNARDELQADKQKWHRRVIEPID--- 76
Db 2968 PGFVLADIDGDEASWDALPRAVA-----SAASEVAIRAGAVVPLARATDEGL 3017
QY 77 -----MDAYRQFLTEIG-----YLLPEPDDFTITTSVGDAEITTTAGPQLVWPVLNARF 125
Db 3018 VVADEAAGPWLVDVTEAGTLANLALVPCPD-----ASRPLGPDEV-----RI 3059
QY 126 ALNANARWGSYDLYGTDVIPETD--GAEKGPYTKV-----RGDKVIAYARKFL 175
Db 3060 AVRAAGV---NFRDYLALGMYPDEGLMGAEAAAGVTVTEVGGVTTLAPGDRVMG----- 3110
QY 176 DSDVPLSSGSGFDATGFTVQDQQLVVALPDKSTGLANPGQFAGYTGAAESTVLLINH 235
Db 3111 -----LVITGGFV---AVTHRMVLRMP-----RCWSFAEAAASVPVAFETAYYA 3152
QY 236 LH-----IEILDPSQVGTDDRAGVKQVI-----LE 262
Db 3153 LHDLAGLGGESVLVHSAAGVGMAAVQARHWDAAEVEFGTASKGKWDVLAAGQDEEHIG 3212
QY 263 SAITTIMPFESVAR-----VDAADKVL--GVENWGLNKDGL---A 299
Db 3213 SSRIT--EFGRFRATSGRGIDVVNLNSGDFVDASARLLREGGR-FVEMKTDITRDL 3269
QY 300 AAVDXDGTAFRLVRNDRNYTAPGG-----GOFTLP-----GR 332
Db 3270 GVVGADGVDPDIRYVAFD--LAEGAERIGQMLDEIMALFDAGVLRPLPPLRAMPVRAHE 3326
QY 333 SLMFV---RNVGH-LMTNDAIVTDGSEVFGIMDALFTGLIAHGLKASDVNGPLNSR 388
Db 3327 ALRFVSOARHVGVKVVLTVPALDAEGTVLITGA--GTLGALVAREHLVTEHDVRRLLVSR 3384
QY 389 TGSIVIVKPKMHGP-----AEVAF--TCELFSR-----VEDVLGLPQNTM-----KIGIM 431
Db 3385 SG-----VAPDLAAELGALGAETVVAACDVANRKAALKALLEDI--PPEHPVTGIVHTAGVL 3438
QY 432 DE-----ERRTTVNLKACIAKAADRVVINTGFLDRTGDEIHT--SWEAGPMVRKG 480
Db 3439 DDGVVSGLTPTPRVDTV-LKPKVDAALTLESVIGELDLPALFVIFSSAASMLGGFGQGSY 3497
QY 481 TMKSQPMILAYEDHNVDAGLAAGFSGRAQVQKGMW-----TWELMADMVETKIAQPRAGA 536
Db 3498 AAANQ-FLDTLARHARRGLTS-----VSLQWGLWHEASGLTGGLADIDRRMS--RAGI 3549
QY 537 STAWVPSPTAATLH-----ALHYQVDVAAVQOGLAGKRATIEQLL----- 578
Db 3550 A-----PMPTDEALHFDRAATELGPVLLPMRLPMRLAALEDRAAD---GTLPELISGLVRVR 3602
QY 579 -----TIP--LAKELAWAPD---BIREVDNCCOSILGYV-----VRWV 612
Db 3603 HRPSARAGTATAAATGPEAFARELAADPPRALRLDLVRGHVALVHSGSPEAIDAEQA 3662
QY 613 DQGVGCSKVPIHDVVALMEDRATLRISQOLLANWRHGVITSAVRAASLERMAPLVDRQN 672
Db 3663 FRDIGFDSLTAVELNRNLNAETGLRLPGLTVFDYPN-----PSALADHLLLELLAP----- 3712

QY 673 AGDVAVRMAPNFDDSIAPLAQELILSGAQOPNGYTEPILHRREREFKARAA 725
Db 3713 ----ATOPTAAPL---LAEERVEOLLGAASPGGPASAVDEBETLITATRIA 3758

RESULT 7
US-10-329-079-45
; Sequence 45, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanouel
; APPLICANT: STAFFA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 45
; TYPE: PRT
; ORGANISM: Streptomyces refuineus
US-10-329-079-45

Query Match 3.5%; Score 132.5; DB 12; Length 5245;
Best Local Similarity 22.0%; Pred. No. 0.12;
Matches 179; Conservative 85; Mismatches 260; Indels 291; Gaps 47;

QY 70 RVIEPIDMDAYQFLTEIGYLLP-----PPDDFTITTSVDIAEITTAGPQLVV 118
Db 4125 RVLEAV-VEAPGTALGEIDVLLPGERELLAGAWSEPPDPGVTTAGAAAD----- 4172

QY 119 PVLNARFALNANARNGSLYDALYGTDPVETDGAKEGPTYNKVR-GDKVIAYARKFLDD 177
Db 4173 -----GIRPPDIFEA-----QAARTHAPAVRDGGEEVAYAE----- 4204

QY 178 SVPLSSGS-----PGDATGFTVQDQLVLPKSTGLANGPGFAGVTGA-----AB 224
Db 4205 ---LNSRANRLARLLAGRGAG---PDTTAVALLP-RGAGLITAVAVQKAGAAVYPLDAE 4257

QY 225 SPTSVLLINHLGHIETLIDPESQVGTTRAGVKDILES---ITTIMD--FEDSVAAYDA 280
Db 4258 LPTG-----RIAHMLDQAKPVLTVLTGMRDALPAGAGPVVCLDDPATRAALAGLDG 4309

QY 281 ADKVLGYRNMGLNKGDLAAAVDKGTAFRLVNLNRDNYTAPGGGQ---FTLPGRSL--M 335
Db 4310 ADCTDADR-----APAGDRDPAYVV-----YTSGSTGTPKGVVVEQSRSLA 4352

QY 336 FVRNVCHLMTNDAIVDTGSEVFEGIMDALFT-----GLIAHGLKASDVNGP----- 383
Db 4353 LVRSAAARYRGAAGTIVLLHGSFAPDLTFTPLVAGGCVIVVADLDAEAGDAPNRPDLK 4412

QY 384 -----LINSRTGSIYIVKPKM-----HGPAEVAFTCE 410
Db 4413 VTPSHLAFLDGIASWAAPADLVVGGEQLTGAPLRLAHPGMRVYNDYGPTEATVSCA 4472

QY 411 LFRSERVEDLGPONTWKIG-----IMDEERTVNLKACIKAAADRVPFI-----N 456
Db 4473 DFV-LFEPGDELPAADVIGRPLAGHRLFVLDRLP-----VPAGVPCELYIAGVGVA 4524

QY 457 TGFILDTGDEIHTSMEA-----GP---MVRKGTMSQPWILAYEDHNVDAGLAAGF 504
Db 4525 RYLGRPG-----MTAREFVCGPFGPGRMVRTGDLAR-W---RPGNLE-----Y 4567

QY 505 SGRVQVKGKMTWTELMDMVKIA---QPRAGAS-----TAWPSPPTAATLHAL 552
Db 4568 LGR---GDGQKVRGFRIEPGEIEAALLDRFEIGQAQVAVLGRBLVAYVAPEA----- 4618

QY 553 HYHQVDVAAVQOGLAGKR-----RATIEQLTTPLA-----KELAWAPDEIR 594
Db 4619 ---BFDPAALREGLAARLPYMWPAATVRLDALPLAPGGKLDHRLPEPPAPADAPHRR 4675

QY 595 EEVDNNQCSILGYVVRWVDQVGS---KVPDIHDVALMEDRATL---RISSQLLANWLR 648
Db 4576 PPRD-----AW---ERVLCFAFRFVLGVAEAGVADDFFALGSDSIGSIQLVGRVR 4722

QY 649 H--GVITSADVRLASLERMAP--LVDRQNAQDVAY-----RPMAPNFDDSDIAFLAAQELIL 699
Db 4723 RAGGRMTVRDV---FERRTPAALAARSQSGAAFEVLGGRATGVPVPTPISSWLAE---L 4776

QY 700 SGACQPNGYTEPILHRREREFKARAAEKAPSDRA 734
Db 4777 GGAEB--GYNOSVLLR-----VPAQADEA 4798

RESULT 8
US-10-080-170-539
; Sequence 539, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 539
; LENGTH: 1624
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-539

Query Match 3.4%; Score 130; DB 16; Length 1624;
Best Local Similarity 20.9%; Pred. No. 0.029;
Matches 135; Conservative 87; Mismatches 216; Indels 208; Gaps 31;

QY 153 AEKG-PTYNKVRGDKVIAYARKFLDDSVPLSSGSFG---DATGFTVQDQLVLPKST 208
Db 916 ADKGTATFSDIANDVAKSYG-FWLGDAP-ASGSGVGYDHRKAMGITAR----- 960

QY 209 GLANPGFAGVTGAESPTSVLLINHLGHI-EIILDPESQ---VGTTRAGVKDV--- 259
Db 961 -----GAWEAVKR-----HFREICIDTQDTFTVVGIDMSG---DVFNG 998

QY 260 -ILESATITIMDFE-----DSVAADAAKDKVLGYRNMGLNKGDLAAAVDKGTAF 310
Db 999 MLKSHIRLIAAFDHRHIFLDPNPDAAVSWAER-----RMFELEPRSSWS----- 1043

QY 311 RVLNDRNYTAPGGQFTLPGRSLMFRNVGHLMTNDIAIVDTGSE-----VFEG 361
Db 1044 ---DYDRSLISEGGVYSREOKAIPLSAQVRAVLGIDGSDVGGAAEMAPPNLIBAILRAP 1100

QY 362 MDALFTGLIATHGLKASDVNGPLNSRTGSIYIVKPKMHPAEVAFTCELSFVEDVLGL 421
Db 1101 VDLFNGIGIY-----IKASESDADV-----DRANDPVRV 1133

QY 422 PQNTMKIGIMDEERTVNLKACIKAAADRVPF-----INTGFLDRTG-----DEIHT 469
Db 1134 NANOVRAKVIPEGGNLGV-----TALGRVEFDLSGGRINTDALDNSAGVDCSDHEVNI 1186

QY 470 SMEAGPMVRKGTMSQPWILAYEDHNVDAGLAAGFSGRAQVKGKMTWTELMDMADWV--ET 527
Db 1187 KILIDSLVSAGTIVKADE-----RTQJ---LESMTDEVAQLVLADN 1223

QY 528 KIAOPRAGASTAWPSPPTAATLHALHVQVDVAAVQOGLAGK-----BRATIEQ 576
Db 1224 EDQNDLNGTSA-----NAASILLPVAHQIKYLVARGVNRELEALPSEKSIARRSEAGI 1278

QY 577 LLTTP-LAKELAWAPDEIRREVDNNQCSILGYVVRWVDQVGS---KVPDIHDVALMEDRAT 635

Db 1279 GUTSPCLATLMAHVKLGLKEEV-----LATELPDQDVAFSLRPLRYPTALRE----- 1325
QY 636 LRISSQLLANWLHGVITSADVRSALERMPLVDRQAGOVAYRPMAPNFDSDSIAPLAQ 695
Db 1326 -RFTPEIRSHQURREIVITMLINDVDTAGITAFRIAEDGVTPF-----DAVRTYVAT 1379
QY 696 ELILSGAQQNGYTEPILHRRERREKAPAAEKP-APSDRAGDDAAR 740
Db 1380 DAIFG-----VGHWR-----IRANLPALSRLTLQTRR 1411

RESULT 9
US-09-905-983-48
; Sequence 48, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Eilat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; CURRENT APPLICATION NUMBER: US/09/905,983
; FILE REFERENCE: 01/2326
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 48
; TYPE: PRT
; LENGTH: 878
; ORGANISM: Homo sapiens
US-09-905-983-48

Query Match 3.4%; Score 128; DB 9; Length 878;
Best Local Similarity 22.0%; Pred. No. 0.016;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY 91 LPEPDDFTI-----TSGVDAEITTTAGPOLVPLNARPALNANARWGSLYDAL 141
Db 307 LPDKMFTINRNTGVISVVTTLGDRSPPTV--TLVW-----QAADLQEGEL--ST 353
QY 142 YGTDVIPETDGAEGPTYNKV--RGD-----KVIAYARKFLDSDVPLSSGSGFGDATGFT 193
Db 354 TATAVITVDNDNPPFNPPTYKGVPEANEVITLKVTDADAPNTPAWEAVTILN 413
QY 194 VQDQQLVVAL--PKSTGLAN-----PGQFAGYTGAAESPTSV 229
Db 414 DGGQFVVTTNPNVNDGILKTAKGLDFEAKQOYILHVAVTNVVPFVSLTSTATVTVDV 473
QY 230 LLINHLHIEILIDPESQVGTDRAGV-----KVILESALTITIMDFEDSVAAV 278
Db 474 LDVNEG---PIFVPEKREVESEDFGVQSEITSYTAQEPDTFMEQKITRYI--WRDT---- 525
QY 279 DAADKVLGYRNWGLN--KGDAAAVDKDGTAFRLVLRNDRNYTAPGGGQFTLPGRSLMF 336
Db 526 -----RWLEINPDTGAISRAELDREDFEHVKN--STVTA----- 559
QY 337 VRNVGHLMTNDALVDTD--GSEVPEGINDALFTGLIAHGLKASDV--NGPLINSRTGSIY 393
Db 560 -----LIATDNGSPVATG-----TGTL--LILSDVNDNAPPEPTIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL--FSRV-----EDVLGLPQNTMKIGIM 431
Db 600 ERNPKQVINIHADLPNTSPPTAEUTHGRVNPWTIQYNDPTQESILKPKMALEVG-- 657
QY 432 DEERTTVNLKACIKAAADRVFNINTGLDRTGDEIHTSMEAGPMVRKMGTKSQPILAY 491
Db 658 ----DYKINUKLMDNQNDQVTTLEVSVCDEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSGRAQVKGKGNWTMTLMADMVETKIAQPRAGASTAWVPSPTAA 547
Db 697 ----VEAGLQIPALGILG-----GILAL--LILILLLLFLRRRAVVKPELLP--PEDD 743
QY 548 TLHALFTH-----QVDVRAVQOGLAGK-----RRATIEQLLTIPIAKELAWAPDEI 593
Db 593

Db 744 TTDNVTYYDEBGGGEDQDFLSQLHRLGDARPEVTRNDVAPTLMSVPRYLPRPANPDEI 803
QY 594 REEVNNQCS 603
Db 804 GNFDENLKA 813

RESULT 10
US-10-165-049-2
; Sequence 2, Application US/10165049
; Publication No. US20020192724A1
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael
; APPLICANT: Cepsek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; FILE REFERENCE: L00560/70010ERP
; CURRENT APPLICATION NUMBER: US/10/165,049
; CURRENT FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: USSN 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: USSN 08/732,429
; PRIOR FILING DATE: 1996-11-01
; PRIOR APPLICATION NUMBER: USSN 09/798,267
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (800)..(808)
; OTHER INFORMATION: HAV tripeptide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2225)..(2295)
; OTHER INFORMATION: transmembrane
US-10-165-049-2

Query Match 3.4%; Score 128; DB 14; Length 878;
Best Local Similarity 22.0%; Pred. No. 0.016;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY 91 LPEPDDFTI-----TSGVDAEITTTAGPOLVPLNARPALNANARWGSLYDAL 141
Db 307 LPDKMFTINRNTGVISVVTTLGDRSPPTV--TLVW-----QAADLQEGEL--ST 353
QY 142 YGTDVIPETDGAEGPTYNKV--RGD-----KVIAYARKFLDSDVPLSSGSGFGDATGFT 193
Db 354 TATAVITVDNDNPPFNPPTYKGVPEANEVITLKVTDADAPNTPAWEAVTILN 413
QY 194 VQDQQLVVAL--PKSTGLAN-----PGQFAGYTGAAESPTSV 229
Db 414 DGGQFVVTTNPNVNDGILKTAKGLDFEAKQOYILHVAVTNVVPFVSLTSTATVTVDV 473
QY 230 LLINHLHIEILIDPESQVGTDRAGV-----KVILESALTITIMDFEDSVAAV 278
Db 474 LDVNEG---PIFVPEKREVESEDFGVQSEITSYTAQEPDTFMEQKITRYI--WRDT---- 525
QY 279 DAADKVLGYRNWGLN--KGDAAAVDKDGTAFRLVLRNDRNYTAPGGGQFTLPGRSLMF 336
Db 526 -----RWLEINPDTGAISRAELDREDFEHVKN--STVTA----- 559
QY 337 VRNVGHLMTNDALVDTD--GSEVPEGINDALFTGLIAHGLKASDV--NGPLINSRTGSIY 393
Db 560 -----LIATDNGSPVATG-----TGTL--LILSDVNDNAPPEPTIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL--FSRV-----EDVLGLPQNTMKIGIM 431
Db 431

Db 600 ERNPKQVINIHADLPNTSPFTAEIHTHGRVNPNTIQYNDPTQESIIILKPKQALEVG-- 657
QY 432 DEERRTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPVKRTKMSQWILAY 491
Db 658 ----DYKINLKLMDNQNKDQVITILEVSVCDCEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSRAQVKGKMTWMTLMADMVETKIAQPRAGASTAWPSPTAA 547
Db 697 ----VEAGLQIPAILGILG-----GIAL--LILILLLLFLRRRAVVKELLP-PEDD 743
QY 548 TLHALHYH-----QVDVAVOOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
Db 744 TRDNVYYDEEGGEEDQDFLSQLHGLDARPEVTRNDVAPILMSVPRYLPRANPDEI 803
QY 594 REEVDNQCOS 603
Db 804 GNFDENLKA 813
RESULT 11
US-10-165-049-3
; Sequence 3, Application US/10165049
; Publication No. US20020192724A1
; GENERAL INFORMATION:
; APPLICANT: Cepek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; TITLE OF INVENTION: Interactions with T Lymphocytes
; FILE REFERENCE: 100560/70010ERP
; CURRENT APPLICATION NUMBER: US/10/165,049
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: USSN 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: USSN 08/732,429
; PRIOR FILING DATE: 1996-11-01
; PRIOR APPLICATION NUMBER: USSN 09/798,267
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-165-049-3

Query Match 3.4%; Score 128; DB 14; Length 878;
Best Local Similarity 22.0%; Pred. No. 0.016;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;
QY 91 LPEPDDFTI-----TTSGVDAEITTTAGPQLVVPVNLARFALNAANARWGSILYDAL 141
Db 307 LPDKNFTINRTGVISVVTGLDRESFTY--TLVV-----QAADLOGEGI--ST 353
QY 142 YGTDVIPETDGAEGKPTYNK--RGD-----KVIAAYAKFLDSDVPLSSGSGFGDATGFT 193
Db 354 TATAVITVTDNDNPPINFTYKGOVPEANVITTLKVTDAADNPAPAEVVTILN 413
QY 194 VQDQQLVAL-PDKSTGLAN-----PCGAGYTGAAESPTSV 229
Db 414 DCGQFVVTNPVNNNGILKTAKGLDPEAKQOYILHVAVTNVVPEVSLTTTATVTDV 473
QY 230 LLINHLGHIILIDPESQVGTTRDAGV-----KDVILESAITTIMDFEDSVAAV 278
Db 474 LDVNEG--PIFVPEKREVESEDFGQGEITSYAQEPDTFMEQKITVRI--WRDT---- 525
QY 279 DAADKVLGVRNWLGLN--KGDIAAADVKGDTAFRLVNLNRDNYTAPGGQFTLPGRSLMF 336
Db 526 -----RNWLEINPDTCALSTRAELDRDREDFEHVKN--STYTA----- 559
QY 337 VRNVGHMTNDALVDTD-GSEVFEGIMDALFTGLIAIHGLKASDV--NGPLINSRTGTY 393

Db 560 -----LIATDNGSPVATG-----TGTL--LILSDVNDNAFIPEPRTIFFC 599
QY 394 IVXPK-----MHG-----PAEVAFTCEL-FSRV-----EDVLGLPONTMKIGIM 431
Db 600 ERNPKQVINIHADLPNTSPFTAEIHTHGRVNPNTIQYNDPTQESIIILKPKQALEVG-- 657
QY 432 DEERRTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPVKRTKMSQWILAY 491
Db 658 ----DYKINLKLMDNQNKDQVITILEVSVCDCEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSRAQVKGKMTWMTLMADMVETKIAQPRAGASTAWPSPTAA 547
Db 697 ----VEAGLQIPAILGILG-----GIAL--LILILLLLFLRRRAVVKELLP-PEDD 743
QY 548 TLHALHYH-----QVDVAVOOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
Db 744 TRDNVYYDEEGGEEDQDFLSQLHGLDARPEVTRNDVAPILMSVPRYLPRANPDEI 803
QY 594 REEVDNQCOS 603
Db 804 GNFDENLKA 813
RESULT 12
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-6
Query Match 3.4%; Score 128; DB 10; Length 11877;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;
QY 9 LRIARVLYDFVNNEALPGTDIDPDSFWAGVDKVVADLTPQNALNARDLQAOIDKWHR 68
Db 10193 LRDRAGVL-----DRVILTGTIEPEFGSGSDGGGADPGAEPEASI---DLDLDAEA---LI 10241
QY 69 RRVIEPID-----MDAYRQPLTEIGYLLPDPDDFTTTTSGVDABEITTTAGPQLVVP 119
Db 10242 RMALGPRTMTSSNEQLVDALRASLKENEEELRKES-----RRRDRRQEPMAIVG 10291
QY 120 VLNARFALNAANARWGSILYDAL-YGTDV---IPETDGAEGKPTYNKVRGDKVIAYAEK-- 173
Db 10292 -MSCRFAGGIRSPB--DLWDAAAGKDLVSEVPPEERGWIDSLYDPFGKGTIVYVNA 10348
QY 174 FLDSVPLSSGSGFG-----DATGFTVQDQQLVVALPDKSTGLANPGQFAGYTGAESPTSV 229
Db 10349 FLDDAAGFADAAFFGISPREALAMPDQROLLEA-----SMEVERAGIDPASV 10396
QY 230 LLINHLGHI-----EILIDPESQVGTTRDAGVXDVILESAITTIMDFEDSVAADA 280
Db 10397 RGTDVGVVGGCYQDYAPDIRVAPEGTGGYV-VTGNSSAVASGSIAYSLGLEGPVAVDT 10455
QY 281 A---DKVLGVRNWLGLNKGDLAAADVKGDTAFRLVNLNRDNYTAPGGQFTLPGRSLMFV 337
Db 10456 ACSSSLVALHLKGLNGDCSTAL-VGGVAVL-----ATPGAFIEFS 10497

```
QY 338 -----RNVGHMTNDAIVDTDGSEV--FEGINDALFTG--LIAIHGKASDVNGPL 384
Db 10498 SQQAAADGRTKGFASAADGLAWGEGVAVLLERLSDRKGRHVLAV-----VRGSA 10550
QY 385 INSTGSIYIVKPKGHGAFAEVAFTCELSRVEDVLGPPQNTMKIGIMDEERRTTVNLKAC 444
Db 10551 INQDGASNGLTAP--HGFSQ-----OHLIRQALAD--AELT-----10582
QY 445 IKAADRVVFINTGFLDRTGDEIHT-----SMEAGPMVRKGTWMSQ-----485
Db 10583 ---SSDUDVVEGHTGTGLDPIEAQALLATYGGCRAPGQPLRLGLTKNSIGHTQAASGV 10639
QY 486 -----PWLAYED--HNVD--AGLA-----AGFSGRA 508
Db 10640 AGVIKMQALRHGVLPKTLHVDEPTQVDNSAGSVELLTEAVDWPERPGRRLRAGVSAGF 10699
QY 509 QVKGKMTWTELMADMVETKIAQPRAGASTA-W-VPSPTAATL-----HALHYHQV 557
Db 10700 VGGTNAHVLEBAPAEVSPAVEPPAGGVVWPVPSAKTSAALDAQIGQLAAVAEDRTDV 10759
QY 558 DVAAVQOGL-----AGKRRATIEQLLTIP--LAKELAWAPDEI-----593
Db 10760 DPAVAARALVDSRTAMEHRAVAVGDSREALRDMPEGLVRGTVTDPGRVAFVFPQGT 10819
QY 594 -----REEVDNN-----COSILGYVVRVVDQGV--GCSKVPDIHDVVALMEDRATL 636
Db 10820 QWAGMGAELDSSPEFAAAMAECECTALSPYVDNSLEAVVRQAPSAPTLDRVDVQVP-VTF 10878
QY 637 RISSOLLANWLHRHGVITSADVRSALRMAPLVDRQAGDVAYRPMAPNFDSD 688
Db 10879 AVMSVLAKVQHHGHTPEAVTCHSOGETA-----AAYVAGALTLLDDA 10920

RESULT 13
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PR1
; ORGANISM: Streptomyces venezuelae
US-09-860-846-6

Query Match 3.4%; Score 128; DB 10; Length 11877;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;

QY 9 LRIARVLVDFVNNALPGCTDIDPDSFWAGVDKVVADLTTPQNALNARDELQACIDKWHR 68
Db 10193 LRDAVL-----DTVLRLTGTIEPEPGSGDGAADFCAGPEAFSI---DOLDRAE---LI 10241
QY 69 RRVTEPID-----MDAYRQFLTEIGVLLPEPDDFTITTSVGDAEITTTAGPQLVVP 119
Db 10242 RVALGPRNTWTSNEQLVDALRSLKENEELRKES-----RRRDRPRPEMAIVG 10291
QY 120 VLNARFALNAANRWGSLYDAL-YGTDV---IPTDGAEXGPTYNKVRGDKVIAYARK-- 173
Db 10292 -MSCRFAGGIRSP--ELWDAAVAGKDLVSEVPERGWDIDSLYDPVFGKRGTTVYRNAA 10348
```

```
QY 174 FLDDSPVSSSFG-----DATGFTVQDQLVVALPDKSTGLANPQGFAGYTGAAESTSV 229
Db 10349 FLDDAAGFAAFFGISPREALAMPQQRQLEA-----SWEVFERAGIDPASV 10396
QY 230 LLINHGLHI-----EILIDPESQVGTTRAGVKDVLILESALTITIMDPEDSVAADA 280
Db 10397 RGTGVTGVVGGCYQYADIRVAPEGTGGYV-VTGNSSAVASGRIAYSLGLEGAIVTVD 10455
QY 281 A---DKVLGYZNWLKGLKGLAAAVDKDGTAFRLVLRNDRNYTAPGGGQFTLPGRSLMFV 337
Db 10456 ACSSSLVALHLKALKGLRNGDCSTAL--VGVVAVL-----ATPGAFIEFS 10497
QY 338 -----RNVGHMTNDAIVDTDGSEV--FEGINDALFTG--LIAIHGKASDVNGPL 384
Db 10498 SQQAAADGRTKGFASAADGLAWGEGVAVLLERLSDRKGRHVLAV-----VRGSA 10550
QY 385 INSTGSIYIVKPKGHGAFAEVAFTCELSRVEDVLGPPQNTMKIGIMDEERRTTVNLKAC 444
Db 10551 INQDGASNGLTAP--HGFSQ-----OHLIRQALAD--AELT-----10582
QY 445 IKAADRVVFINTGFLDRTGDEIHT-----SMEAGPMVRKGTWMSQ-----485
Db 10583 ---SSDUDVVEGHTGTGLDPIEAQALLATYGGCRAPGQPLRLGLTKNSIGHTQAASGV 10639
QY 486 -----PWLAYED--HNVD--AGLA-----AGFSGRA 508
Db 10640 AGVIKMQALRHGVLPKTLHVDEPTQVDNSAGSVELLTEAVDWPERPGRRLRAGVSAGF 10699
QY 509 QVKGKMTWTELMADMVETKIAQPRAGASTA-W-VPSPTAATL-----HALHYHQV 557
Db 10700 VGGTNAHVLEBAPAEVSPAVEPPAGGVVWPVPSAKTSAALDAQIGQLAAVAEDRTDV 10759
QY 558 DVAAVQOGL-----AGKRRATIEQLLTIP--LAKELAWAPDEI-----593
Db 10760 DPAVAARALVDSRTAMEHRAVAVGDSREALRDMPEGLVRGTVTDPGRVAFVFPQGT 10819
QY 594 -----REEVDNN-----COSILGYVVRVVDQGV--GCSKVPDIHDVVALMEDRATL 636
Db 10820 QWAGMGAELDSSPEFAAAMAECECTALSPYVDNSLEAVVRQAPSAPTLDRVDVQVP-VTF 10878
QY 637 RISSOLLANWLHRHGVITSADVRSALRMAPLVDRQAGDVAYRPMAPNFDSD 688
Db 10879 AVMSVLAKVQHHGHTPEAVTCHSOGETA-----AAYVAGALTLLDDA 10920

RESULT 14
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PR1
; ORGANISM: Streptomyces venezuelae
US-09-836-821-6

Query Match 3.4%; Score 128; DB 11; Length 11877;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;

QY 9 LRIARVLVDFVNNALPGCTDIDPDSFWAGVDKVVADLTTPQNALNARDELQACIDKWHR 68
```

```
Db 10193 LRDAGVL-----DIVLRLTGIEPEPGSGSDGGAADPGAEPEASI---DDLDAEA---LI 10241
QY 69 RRVIEPID-----MDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVVP 119
Db 10242 RMLGPRNTMTSSNEQLVDALRASLKENEELRKES-----RRRADRCEPWAIVG 10291
QY 120 VLNAREFALNANARWGLSDAL-YGTDV---IPETDGAEGKPTYNKVRGDKVIAYARK-- 173
Db 10292 -MSCRFAGGIRSP--DLMDAVAAGKDLVSEVPEERGWIDSLYDPVGRKGTYYVRNAA 10348
QY 174 FLDSVPLSGSGFG---DATGFTVQDGLVVALPKSTGLANPGQFAGYTGAAESPTSV 229
Db 10349 FLDDAAGFADAFGISPREALAMPQORQLLEA-----SWEFERAGIDPASV 10396
QY 230 LLINHGHLI-----ELIDPESQVGTDRAGVKDVILESAITTIMDFEDSVAADA 280
Db 10397 RGTDVGVVCGGYQDYAPDIRVAPEGTGGYV-VTGNSSAVASGRIAYSLGLEGPAVTVD 10455
QY 281 A---DKVLGYRNWGLNKGDLAAAVDKGTAFRLVLRNDRNYTAPGGGQFTLPGRSLMFV 337
Db 10456 ACSSSLVALHKLKGLNGDCSTAL-VGGVAVL-----ATPGAFIEFS 10497
QY 338 -----RNVGHLMTNDAIVDTDGSEV---FEGIMDALFTG--LIAIHGLKASDVNGPL 384
Db 10498 SQQAAADGRTKGFASAADGLANGEGVAVLLERLSDARRKGRVLAV-----VRGSA 10550
QY 385 INSRTGSIYIVKPMHGPAAVEAFTCELSRVEDVLGLPONTWKIGIMDEERTTNVLKAC 444
Db 10551 INQDASNGLTAP--HGFSQ-----OHLIRQALAD--ARLT----- 10582
QY 445 IKAADRVVFINFGFLDRTGDEIHT-----SMEAGPMVRKGMKSO----- 485
Db 10583 ---SSDVDVVEGHGTGRLGDPTEAQALLATYQGRAPQPLRLGLTKSNIGHTQAAAGV 10639
QY 486 -----PWILAYED--HNVD--AGLA-----AGFSGRA 508
Db 10640 AGVIKVMQALRHGVLPKTLHVDEPTQDQVDSAGSVELLTEAVDWPFRPGLRRAGVSAGF 10699
QY 509 QVKGGMWMTMELMADWVETKIAOPRAGASTA-W-VPSPTAATL-----HALHYHQV 557
Db 10700 VGGTNAHVLEEAFAVEESPAGGVPVWPVSAKTSAALDAQIGQLAAAYAEEDTDV 10759
QY 558 DVAAVQOGL-----AGKRATIEQLLTIP--LAKELAWAPDEI----- 593
Db 10760 DPAVAARALVDSRTAMEHRAVAGDSREALRDMPEGLVRGTVDPCRVAFFVFGQGT 10819
QY 594 -----REEVDNN-----COSILGYVVRVWDQGV--GCSKVPDIHDVVALMEDRATL 636
Db 10820 QWAGMGAEILDSPEFAAANAECEALSPYVDKSLAEVVRQAPSAPFTLDRVDVQVP-VTF 10878
QY 637 RISSQLLANWLRHGVTITSADVRASLERMAPLVDQRQVAGDVAYRPMAPNFDDS 688
Db 10879 AVMVSLAKVQHHGHTPEAVIGHSQGEIA-----AAVAGALTLLDDA 10920
```

RESULT 15

```
US-10-271-889-49
; Sequence 49, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
; FILE REFERENCE: 600 582US1
; CURRENT APPLICATION NUMBER: US/10/271,889
; PRIOR FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
```

Query Match 3.4%; Score 128; DB 12; Length 11877;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;

QY 9 LRIARLVYDFVNNALPGTDIDPDPSFWAGVDVADVADLTPOQALLNARDELQAOIDKWHR 68
Db 10193 LRDAGVL-----DIVLRLTGIEPEPGSGSDGGAADPGAEPEASI---DDLDAEA---LI 10241
QY 69 RRVIEPID-----MDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVVP 119
Db 10242 RMLGPRNTMTSSNEQLVDALRASLKENEELRKES-----RRRADRCEPWAIVG 10291
QY 120 VLNAREFALNANARWGLSDAL-YGTDV---IPETDGAEGKPTYNKVRGDKVIAYARK-- 173
Db 10292 -MSCRFAGGIRSP--DLMDAVAAGKDLVSEVPEERGWIDSLYDPVGRKGTYYVRNAA 10348
QY 174 FLDSVPLSGSGFG---DATGFTVQDGLVVALPKSTGLANPGQFAGYTGAAESPTSV 229
Db 10349 FLDDAAGFADAFGISPREALAMPQORQLLEA-----SWEFERAGIDPASV 10396
QY 230 LLINHGHLI-----ELIDPESQVGTDRAGVKDVILESAITTIMDFEDSVAADA 280
Db 10397 RGTDVGVVCGGYQDYAPDIRVAPEGTGGYV-VTGNSSAVASGRIAYSLGLEGPAVTVD 10455
QY 281 A---DKVLGYRNWGLNKGDLAAAVDKGTAFRLVLRNDRNYTAPGGGQFTLPGRSLMFV 337
Db 10456 ACSSSLVALHKLKGLNGDCSTAL-VGGVAVL-----ATPGAFIEFS 10497
QY 338 -----RNVGHLMTNDAIVDTDGSEV---FEGIMDALFTG--LIAIHGLKASDVNGPL 384
Db 10498 SQQAAADGRTKGFASAADGLANGEGVAVLLERLSDARRKGRVLAV-----VRGSA 10550
QY 385 INSRTGSIYIVKPMHGPAAVEAFTCELSRVEDVLGLPONTWKIGIMDEERTTNVLKAC 444
Db 10551 INQDASNGLTAP--HGFSQ-----OHLIRQALAD--ARLT----- 10582
QY 445 IKAADRVVFINFGFLDRTGDEIHT-----SMEAGPMVRKGMKSO----- 485
Db 10583 ---SSDVDVVEGHGTGRLGDPTEAQALLATYQGRAPQPLRLGLTKSNIGHTQAAAGV 10639
QY 486 -----PWILAYED--HNVD--AGLA-----AGFSGRA 508
Db 10640 AGVIKVMQALRHGVLPKTLHVDEPTQDQVDSAGSVELLTEAVDWPFRPGLRRAGVSAGF 10699
QY 509 QVKGGMWMTMELMADWVETKIAOPRAGASTA-W-VPSPTAATL-----HALHYHQV 557
Db 10700 VGGTNAHVLEEAFAVEESPAGGVPVWPVSAKTSAALDAQIGQLAAAYAEEDTDV 10759
QY 558 DVAAVQOGL-----AGKRATIEQLLTIP--LAKELAWAPDEI----- 593
Db 10760 DPAVAARALVDSRTAMEHRAVAGDSREALRDMPEGLVRGTVDPCRVAFFVFGQGT 10819
QY 594 -----REEVDNN-----COSILGYVVRVWDQGV--GCSKVPDIHDVVALMEDRATL 636
Db 10820 QWAGMGAEILDSPEFAAANAECEALSPYVDKSLAEVVRQAPSAPFTLDRVDVQVP-VTF 10878
QY 637 RISSQLLANWLRHGVTITSADVRASLERMAPLVDQRQVAGDVAYRPMAPNFDDS 688
Db 10879 AVMVSLAKVQHHGHTPEAVIGHSQGEIA-----AAVAGALTLLDDA 10920

Mon Nov 24 15:01:42 2003

us-09-688-672a-2.rapb

Page 10

Search completed: November 21, 2003, 16:38:09
Job time : 32.8617 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 15.4376 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-2

Perfect score: 3810

Sequence: 1 TDRVSVGNLRIARVLYDFVN.....KARAAEKPAAPSRAAGDRAAR 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3810	100.0	741	2 F70722	probable glcB prot
2	3117.5	81.8	731	2 T44752	probable malate sy
3	2569	67.4	725	2 H83586	malate synthase G
4	2374.5	62.3	744	2 AH2582	malate synthase G
5	2374.5	62.3	744	2 F97364	malate synthase G
6	2318	60.8	728	2 AF3299	malate synthase (E
7	2289	60.1	739	2 I40715	malate synthase (E
8	2256.5	59.2	727	2 E83916	malate synthase BH
9	2154.5	56.5	723	2 S51788	malate synthase (E
10	205.5	5.4	542	2 S17774	malate synthase (E
11	190.5	5.0	533	2 D91245	malate synthase A
12	190.5	5.0	533	2 B86093	malate synthase A
13	190	5.0	533	2 AD1011	malate synthase (E
14	189.5	5.0	533	1 SYEQMA	malate synthase (E
15	187	4.9	556	2 H82287	malate synthase A
16	186	4.9	551	2 JX0196	malate synthase (E
17	185	4.9	551	2 JX0196	malate synthase (E
18	175.5	4.6	525	2 A87468	malate synthase (E
19	175	4.6	528	2 S17773	malate synthase (E
20	171.5	4.5	559	2 T03412	malate synthase (E
21	165	4.3	554	2 S26645	malate synthase (E
22	165	4.3	554	2 S48493	malate synthase (E
23	160	4.2	555	1 SYHQMA	malate synthase (E
24	160	4.2	562	2 T48413	malate synthase-11
25	154.5	4.1	562	2 E82396	malate synthase-re
26	151.5	4.0	564	2 T07690	malate synthase (E
27	150.5	4.0	561	1 SYRPM	malate synthase (E
28	149.5	3.9	566	2 S44186	malate synthase (E
29	149	3.9	532	2 AF0453	malate synthase (E

30	149	3.9	566	2 S15387	malate synthase (E
31	144.5	3.8	568	1 SYKWA	malate synthase (E
32	139	3.6	3455	2 B92519	hemagglutinin-like
33	138	3.6	567	1 SYCSM2	malate synthase (E
34	137.5	3.6	4848	2 T30289	pristinamycin I sy
35	135.5	3.6	521	2 D75581	malate synthase -
36	135.5	3.6	3442	2 B82589	hemagglutinin-like
37	134.5	3.5	1441	2 B86807	hypothetical prote
38	130	3.4	1624	2 C70867	probable Helix-tur
39	129.5	3.4	597	2 A90770	hydrogenase-1 larg
40	129.5	3.4	597	2 B85632	hydrogenase-1 larg
41	129	3.4	5188	2 B85547	probable RTX famli
42	129	3.4	5291	2 F90696	hypothetical prote
43	128.5	3.4	597	1 HQECL	hydrogenase (EC 1.
44	128.5	3.4	808	2 F70720	probable transloca
45	127.5	3.3	26926	1 I38344	titin, cardiac mus

ALIGNMENTS

RESULT 1

F70722

probable glcB protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: F70722

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S. Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70722

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-741 <COL>

A:Cross-references: GB:278020; GB:AL123456; NID:93261625; PIDN:CAB01465.1; PID:e257679;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: glcB

Query Match	100.0%	Score 3810	DB 2	Length 741
Best Local Similarity	100.0%	Pred. No. 3.7e-246	Mismatches 0	Indels 0
Matches 740	Conservative 0			
QY	1	TDRSVGNLRIARVLYDFVNEALFGTDIDPDSFWAGYDKVVAADLTQNOALLNARDLQ	60	
Db	2	TDRSVGNLRIARVLYDFVNEALFGTDIDPDSFWAGYDKVVAADLTQNOALLNARDLQ	61	
QY	61	AQIDKWHRRVIEIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAITTTAGPOLVVPV	120	
Db	62	AQIDKWHRRVIEIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAITTTAGPOLVVPV	121	
QY	121	LNARFALNANARNGSLYDALYGTDPVETDGAEGKGTYNKVRGDKVIAYARKELDSDVP	180	
Db	122	LNARFALNANARNGSLYDALYGTDPVETDGAEGKGTYNKVRGDKVIAYARKELDSDVP	181	
QY	181	LSSGSFGDATGFTVDGQGVVALPKSTGLANPQFGAGYTGAAESPTSVLLINHLHIEI	240	
Db	182	LSSGSFGDATGFTVDGQGVVALPKSTGLANPQFGAGYTGAAESPTSVLLINHLHIEI	241	
QY	241	LIDPESQVGTTRAGVKDVILESAITTIMDPEDSVAADKVLGVRNVLGNKGLDAA	300	
Db	242	LIDPESQVGTTRAGVKDVILESAITTIMDPEDSVAADKVLGVRNVLGNKGLDAA	301	
QY	301	AVDXDGTAFRLVNRDNYTAPGGQFTLPGRSLMFRVNVGHLMTNDAIVDTDGSVEFEG	360	
Db	302	AVDXDGTAFRLVNRDNYTAPGGQFTLPGRSLMFRVNVGHLMTNDAIVDTDGSVEFEG	361	
QY	361	IMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPMHGPAFVATCELFSRVEDVLG	420	

Db 362 IMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSFSEVEDVLG 421
Qy 421 LPONTMKGIMDEBERTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPVVRKG 480
Db 422 LPONTMKGIMDEBERTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPVVRKG 481
Qy 481 TMKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADVMETKIAQPRAGASTAW 540
Db 482 TMKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADVMETKIAQPRAGASTAW 541
Qy 541 VPSSTAAATLHALHYHVDVAQVQGLAGKRRATIEQLTTIPLAKELAWADEIREEDNN 600
Db 542 VPSSTAAATLHALHYHVDVAQVQGLAGKRRATIEQLTTIPLAKELAWADEIREEDNN 601
Qy 601 COSILGYVVRWVDQGVGCSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITSADVRAS 660
Db 602 COSILGYVVRWVDQGVGCSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITSADVRAS 661
Qy 661 LERMAPLVDROAGDVAYRPMAPNFDDSIAPLAAQELILSGAQCPNGYTEPIHLRRREF 720
Db 662 LERMAPLVDROAGDVAYRPMAPNFDDSIAPLAAQELILSGAQCPNGYTEPIHLRRREF 721
Qy 721 KARAAEPAPSDRAGDDAAR 740
Db 722 KARAAEPAPSDRAGDDAAR 741

RESULT 2
T44752
probable malate synthase (EC 4.1.3.2) G [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 05-May-2000
C:Accession: T44752
R:Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z22833
A:Accession: T44752
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-731 <PAR>
A:Cross-references: EMBL:AL008609; PIDN:CAA15459.1
A:Experimental source: cosmid B1788
C:Genetics:
A:Note: glcB
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 81.8%; Score 3117.5; DB 2; Length 731;
Best Local Similarity 82.0%; Pred. No. 5.7e-200;
Matches 597; Conservative 54; Mismatches 72; Indels 5; Gaps 2;
Qy 1 TDRVSAGNLRLARVLYDFVNEALPGTDIDPDPSWAGVDKVVADLTPOQALLNARDLQ 60
Db 2 TDRVSAGNLRLARVLYDFVNEALPGTDIDPDPSWAGVDKVVADLTPOQALLNARDLQ 61
Qy 61 AQIDKWHRRVIEPIDMDAYQFTTEIGLYLLPEPDDFTITTSVDAEITTTAGPQLVVPV 120
Db 62 AQIDKWHRRVIEPIDMDAYQFTTEIGLYLLPEPDDFTITTSVDAEITTTAGPQLVVPV 121
Qy 121 LNARFALNANARWGLSDYALYGTDTIPEETGASKEGYNKIRGDKVIAYARKFLDSDVP 180
Db 122 LNARFALNANARWGLSDYALYGTDTIPEETGASKEGYNKIRGDKVIAYARKFLDSDVP 181
Qy 181 LSSGSFGDAGTFTVQDGLVVALPKSTGLANPQFAGYTGAAESPT-SVLLINHLHIE 239
Db 182 LASDSWTNATGVSIPDGLQIAIGNSTGLASPEKFGVYNRQLRSSNWSVLLANHLHIE 241
Qy 240 ILIDPESQVGTDRAGVDKLVILESAITTIMDFESVAAVDAADKVLGVYRWLGNKGLA 299
Db 242 VLIDPESPIGKTPVGIKNDVILESAITTIMDFESVAAVDAADKVLGVYRWLGNKGLD 301
Qy 300 AAVDKDGTAFRLNDRNRYTAPGGQFTPLCRSLMFPVRYNCHLMTNDAY-...DTDGS 355
Db 302 BEVNKGDTFTFRLNADRSYTPDGGELTLPGRSLLFVRNCHLMTTSDAILVDGDDGQEK 361

Qy 356 EVFEGIMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSF 415
Db 362 EVFEGIIDAVFTGLAAIHGLTGEANGPLTNSRTGSIYIVKPKMHGPAEVAFTCELSF 421
Qy 416 EDVLGLPONTMKGIMDEBERTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGP 475
Db 422 EDVLGLPONTMKGIMDEBERTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGP 481
Qy 476 MYRKGTMTKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADVMETKIAQPRAG 535
Db 482 MYRKGTMTKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADVMETKIAQPRAG 541
Qy 536 ASTAWVPSTAAATLHALHYHVDVAQVQGLAGKRRATIEQLTTIPLAKELAWADEIRE 595
Db 542 ASTAWVPSTAAATLHALHYHVDVAQVQGLAGKRRATIEQLTTIPLAKELAWADEIRE 601
Qy 596 EVDNCCQILGYVVRWVDQGVGCSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITS 655
Db 602 EVDNCCQILGYVVRWVDQGVGCSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITS 661
Qy 656 DVRASLERMAPLVDROAGDVAYRPMAPNFDDSIAPLAAQELILSGAQCPNGYTEPIHLR 715
Db 662 DVRASLERMAPLVDROAGDVAYRPMAPNFDDSIAPLAAQELILSGAQCPNGYTEPIHLR 721
Qy 716 RRRREFKAR 723
Db 722 RRRREFKAR 729

RESULT 3
H83586
malate synthase G PA0482 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83586
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lim,
.: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: H83586
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <STO>
A:Cross-references: GB:AE004485; GB:AE004091; NID:G9946332; PIDN:AAG03871.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: glcB; PA0482
Query Match 67.4%; Score 2569; DB 2; Length 725;
Best Local Similarity 69.1%; Pred. No. 2.2e-163;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;
Qy 1 TDRVSAGNLRLARVLYDFVNEALPGTDIDPDPSWAGVDKVVADLTPOQALLNARDLQ 60
Db 2 TDRVSAGNLRLARVLYDFVNEALPGTDIDPDPSWAGVDKVVADLTPOQALLNARDLQ 61
Qy 61 AQIDKWHRRVIEPIDMDAYQFTTEIGLYLLPEPDDFTITTSVDAEITTTAGPQLVVPV 120
Db 62 AQIDKWHRRVIEPIDMDAYQFTTEIGLYLLPEPDDFTITTSVDAEITTTAGPQLVVPV 121
Qy 121 LNARFALNANARWGLSDYALYGTDTIPEETGASKEGYNKIRGDKVIAYARKFLDSDVP 180
Db 122 LNARFALNANARWGLSDYALYGTDTIPEETGASKEGYNKIRGDKVIAYARKFLDSDVP 181
Qy 181 LSSGSFGDAGTFTVQDGLVVALPKS-TGLANPQFAGYTGAAESPT-SVLLINHLHIE 239
Db 182 LSSGSFGDAGTFTVQDGLVVALPKS-TGLANPQFAGYTGAAESPT-SVLLINHLHIE 241
Qy 240 ILIDPESQVGTDRAGVDKLVILESAITTIMDFESVAAVDAADKVLGVYRWLGNKGLA 299

Db 242 IOTDPPSPGQTDAGAKVDVMEALTTIMDCEDSVAAVDADDKVVIVRNWLGMLKMGDLA 301

QY 300 AAVDKDGTAFVLARDRNYTAPGGQFTLPGRSLMEVENVGHMTNDALVDTDCSEVPE 359

Db 302 EVSKGGSTFTTMDPDRVYTTADGSELTLHGRSLTFVRNVGHMTNDAILDKDGNVPE 361

QY 360 GIMDALFTGLIAIHGLKASDVNGPLI--NSRTGSIYIVKPKVHGPAEVAFTCELSERVED 417

Db 362 GICDGLFTSLIAIH-----DLNGNTRSKNSRTGSIYIVKPKVHGPEEAAFTNELFGRVED 416

QY 418 VLGLPONTMKGIMDEBERTTNVNLKACIAAADRNVFINTGLDRTGBEIHSTMEAGPMV 477

Db 417 VLGLPNTLUKVGIMDEBERTTNVNLKACIAAADRNVFINTGLDRTGBEIHSTMEAGAV 476

QY 478 RKGTMKSQPVILAYEDHNVDAAGLAFSGRAQVGVKGMWMTMELMADMVETKIAQPRAGAS 537

Db 477 RKGAMKSEKWI GAYENNVDVGLATGLQKQAQIGKGMWMPDLMAMLEQKIGHPLAGAN 536

QY 538 TAWVPSPTAATHALHYHOVDVAAVQOGLAGKRRATIEBOLLTIPLAKELAWAPDIREV 597

Db 537 TAWVPSPTAATHALHYHKVDVFAQAEAKRTPASVDLITPLAPNTNMTAEBIKREV 596

QY 598 DNNQCSILGVVVRWDQVGVCSKVPDIHDVVALMEDRATLRISQLLANLWLRHGVITSADV 657

Db 597 DNNAGILGVVVRWDQVGVCSKVPDINDVGLMEDRATLRISQLLANLWLRHGVISQEV 656

QY 658 RASLERMAPLVRQNAVGVAYRPMAPNFDDSIATFAAQELILSGAQPNGYTEPILHRRR 717

Db 657 VESLKMNAVVRQNASDPSYRPMAPNFDDNVAFQAALVELVVEGTRQPNGYTEPVLHRRR 716

QY 718 REFKA 723

Db 717 REFKA 722

RESULT 4

AH2582

malate synthase G [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C:Accession: AH2582

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.; Wood, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:21608550; PMID:11743193

A:Accession: AH2582

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-744 <KUR>

A:Cross-references: GB:AE008688; PIN:AAL41078.1; PID:g17738367; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: glcB

A:Map position: circular chromosome

Query Match 62.3%; Score 2374.5; DB 2; Length 744;

Best Local Similarity 62.3%; Pred. No. 2.1e-150;

Matches 462; Conservative 97; Mismatches 162; Indels 21; Gaps 4;

QY 1 TDRVSVGNL-----RIARV-----LYDFVNEALPGTIDPDPSFWAGVKVA 43

Db 6 SRLSVQSPSEYKEAHVSRDTKFLSIDRLIYAFLTEVLPTGLDSEITFEFGSAIVH 65

QY 44 DLTPOQALLNARDELQADIKWHRRIEIPDMDAYQFLTEIGYLLPEDDFTITSG 103

Db 66 ELSKPNRELLAKDALQEKIDGWRENG-APSDFDAYEAFLEKEIGYLLPEGGFGFKVETNN 124

QY 104 VDAEITTTAGPOLVPLNARALNANRWGLSDYALYGTDIVIETPGAEGKPTYNKR 163

Db 125 VDPEIAVAGPQLVVPVMNARYALNANRWGLSDYALYGTDAISDADGAEGKRGYNPKR 184

QY 164 GDKVIAARFELDDSVPLSSGFDATFTVQDQQLVVALPDKSTGLANPQAGYTGAA 223

Db 185 GDKVIAWARFLDESAPLETGSDVDTGFNIADGLQLAI GAATTLGLKDAVQVQFGSGEA 244

QY 224 ESPTSVLLIHLHIEILIDPESOVGTTDRAGVKDVILESAITTIMDFESVAAVDAADK 283

Db 245 AKPATILLKGNLHTEIVDPSTEIGKSDRAGISDVILESALTITMDCESVAAVDAEDK 304

QY 284 VLGVRNWLGKNGDLAAAVDKGTAFURVLRNRDNYTAPGGQFTLPGRSLMFVRNVGHL 343

Db 305 VLVTGNVLGLMRGDLTEAVSGKGNFTFRLNPRDYYTAPDGSALTLPGRSLMLVRNVGHL 364

QY 344 MTNDALVDTDGSSEVFEGIMDALFTGLTAIHGLKASDVNGPLINSRTGSIYIVKPKMHGPA 403

Db 365 MTNPAILDRGDRVPEGIMDAVVTALLALYDVGPS---GRQNSRAGSMVIVKPKMHGPE 421

QY 404 EVAFTCLFSEVEDLGLPQNTMKIGIMDEBERTTNVNLKACIAAADRNVFINTGLDRT 463

Db 422 EVAFANEIFARVENLVGMAFNTMKGIMDEBERTTNVNLKESIRAAKDRVVVINTGLDRT 481

QY 464 GDEHTSWEAGPMVKCTMKSQPVILAYEDHNVDAAGLAFSGRAQVGVKGMWMTMELMAD 523

Db 482 GDEHTSWEAGPMIRKGMKQAAIAAYENNVNDIGLECSLGHQAQIGKGMWMPDLMAA 541

QY 524 MVETKIAQPRAGASTAVVPSPTAATHALHYHOVDVAAVQOGLAGKRRATIEQLTIPLA 583

Db 542 MLEQKIAHPKAGANTAVVPSPTAATHALHYHKVDVAAVQOGLKSRGAKLSDLISVPVA 601

QY 584 KELAWAPDEETREEDVNNQCSILGVVVRWDQVGVCSKVPDIHDVVALMEDRATLRISQLL 643

Db 602 PRPNWTPETIQRLELDNNAQILGVVVRWDQVGVCSKVPDINNIGLMEDETRATLRISAQM 661

QY 644 ANWLHRGVITSADVRASLERMAPLVRQNAVGVAYRPMAPNFDDSIATFAAQELILSGAQ 703

Db 662 ANWLHRGVITAEQIILTKMKMAAVVDTONAGDPAFLPMASDFDGSVAFQAVALVELKGRE 721

QY 704 QPNGYTEPILHRRRREFKARAA 725

Db 722 QPNGYTEPILHRRRLLEKAKQA 743

RESULT 5

F97364

malate synthase G (P04082) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002

C:Accession: F97364

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens Cereon

A:Reference number: A97359; PMID:21608551; PMID:11743194

A:Accession: F97364

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-744 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK85871.1; PID:g15154912; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C78

A:Map position: circular chromosome

Query Match 62.3%; Score 2374.5; DB 2; Length 744;

Best Local Similarity 62.3%; Pred. No. 2.1e-150;

Matches 462; Conservative 97; Mismatches 162; Indels 21; Gaps 4;

QY 1 TDRVSVGNL-----RIARV-----LYDFVNEALPGTIDPDPSFWAGVKVA 43

Db 6 SRLSVQSPSEYKEAHVSRDTKFLSIDRLIYAFLTEVLPTGLDSEITFEFGSAIVH 65

QY 44 DLTPOQALLNARDELQADIKWHRRIEIPDMDAYQFLTEIGYLLPEDDFTITSG 103

Db 66 ELSKPNRELLAKDALQEKIDGWRENG-APSDFDAYEAFLEKEIGYLLPEGGFGFKVETNN 124

QY 104 VDAEITTTAGPOLVPLNARALNANRWGLSDYALYGTDIVIETPGAEGKPTYNKR 163

Db 66 ELSPQNRRELLAKRDLQEKIDGWYRENG-APSPDFDAYEAFLEKIGYLLPEGPGFKVETNN 124
Qy 104 VDAEITTTAGPQLVVPVNLNARFALNANARWGSYDALYGTVDVTPETDGAEGKPTYNKVR 163
Db 125 VDPEIAVAGPQLVVPVNNARYALNANARWGSYDALYGTDAISADGAEGKGYNPKR 184
Qy 164 GDKVIARPLDSDVPLSSGSGDGTFTVQDQGLVVALPDKSTGLANGPGQAGYTGAA 223
Db 185 GDKVIARWFLDSDAPLETGWSVDTFGNFIADQLQLAIGATTTGLKDAVQKFGSGEA 244
Qy 224 ESPTSLLIHGHGHIETILIDPESQVGTTRAGVKDVIKESAITTINDFDSVAADK 283
Db 245 AKPATILGKNGULHETIIVDPSTIEGKSDRAGISDVIKESAITTINDFDSVAADK 304
Qy 284 VLGVRNMLGNKGLAAADKDTAFULVRNDRNTYAPGGGFTLPGRSIMFVRNVGHL 343
Db 305 VLVGNNMLGMRGLTEAVSKGNTFTFRLNDRNTYAPDGSALTLPGRSLMLVRNVGHL 364
Qy 344 MTNDALVDTGSEVFEIGIMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMHGPA 403
Db 365 MTNPAILDRDGRDVEGIMDAVVTALIALYDVGPS---GRQNSRAGSMYVVRKWHGPE 421
Qy 404 EVAFTCLFSVEDVLGPQNTMKIGIMDEERTTNLAKACIAAADRVVVFINTGFLDRT 463
Db 422 EVAFAEIRFARVENLVGMATNWKMGIMDEERTTNLAKESIRAAKDRVVVFINTGFLDRT 481
Qy 464 GDEIHTSMEAGPVRKGTMSQPMILAYEDHNVDAAGLAFSGRAQVKGKMTWELMAD 523
Db 482 GDEIHTSMEAGPMTKMDKQAAWIAAYENNVNDIGLEGLSGHAQIGKGMWAMPILMAA 541
Qy 524 MVEYKIAOPRAGASTAVPSPSTAATHALYHVDVAAGVQGLAGKRATIEOLLTIPLA 583
Db 542 MLEOKIAHPKACANTAVPSPSTAATHATHYKVDVAAGVQGLAGKRATIEOLLTIPLA 601
Qy 584 KELAWAPDEIRREVDNNQCSILGVYVRWDQGVCSKVPDIHVALMEDRATIRISQLL 643
Db 602 PRPNWTPEEIQRELDNNAQSIGLYVVRWDQGVCSKVPDINNIGLMEDRATIRISAQHM 661
Qy 644 ANWLHRGVITSADYRASIERNAPLVRQNDGVAIRPMAPNFDSDIAFLAAQELISGAQ 703
Db 662 ANWLHRGVITBAQIKTKRNVAAVDTQNDGDPAYLPMASDFDGSVAFQAAVELVLKRE 721
Qy 704 QPNGYTPILHRRRRREFKARAA 725
Db 722 QPNGYTPVLHRRRLLELKAKQA 743

RESULT 6

AF3299
malate synthase (EC 4.1.3.2) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AF3299
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Ietess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3299
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-728 <KUR>
A:Cross-references: GB:AB008917; PIDN:AAU51561.1; PID:gl7982281; GSFDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0380
A:Map position: 1
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 60.8%; Score 2318; DB 2; Length 728;
Best Local Similarity 61.8%; Pred. No. 1.2e-146;
Matches 444; Conservative 97; Mismatches 174; Indels 4; Gaps 2;

Qy 4 VSVGNLRIRARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTLPONQALLNARDELQAOI 63
Db 10 VEIEGLAVAPLEVEFLAKEAPGTGVEPEKFWKFAAIIRDLPKRNALLAKRDELQARI 69
Qy 64 DKWHRRRVIEPDMDAYRQFLTEIGYLLPPDDFTITTSVGDABEITTTAGPQLVVPVYLA 123
Db 70 DAWKENRDKYSQADVQOFLKIDIGYLLPEGASVSTTNVDPEITHIAGPQLVVPVYLA 129
Qy 124 RFALNANARWGSYDALYGTVDVTPETDGAEGKPTYNKVRGDKVIAVARFLDSDVPLSS 183
Db 130 RYALNANARWGSYDALYGTDAISEADGAEGKGYNPKRGEKVIKAWKNFLDSEAPLST 189
Qy 184 GSGDGTCTFTVQDQGLVVALPDKS-TGLANPGQAGYTGAAESPTSVLLNHGHGHIILI 242
Db 190 GKWADVAGLANVKGLEIRLTGDSATLLKDESQFKGYNGDAASPTNVLLAKHNVHVDIV 249
Qy 243 DPESQVGTTRAGVKDVIKESAITTINDFDSVAADKAVLGYRNWGLNKGDLAAAV 302
Db 250 NADHPIGKTDPAHIADVVLESASITIDCEDSIADVAEDKAVYRNWGLNKGKLEDTF 309
Qy 303 DKDGTAFULVRNDRNTYAPGGGFTLPGRSIMFVRNVGHLMTNDALVDTGSEVFEIGIM 362
Db 310 EKNGQMTFRLNGDRTYAPDGSALTLPGRSLMLVRNVGHLMTNPAILDAGNEVPEGIM 369
Qy 363 DALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVLGLP 422
Db 370 DAAFTSLIALHDI---GNGRHMNSREGSVYIVKPKHGPPEVAFANEFTFTEMLGMK 426
Qy 423 QNTMKIGIMDEERTTNLAKACIAAADRVVVFINTGFLDRTGDEIHTSMEAGPVRKGTM 482
Db 427 PNTLIGIMDEERTTNLAKESIRAAKDRVVVFINTGFLDRTGDEIHTSMEAGPVRKGTM 486
Qy 483 KSQPMILAYEDHNVDAAGLAFSGRAQVKGKMTWELMADMTETKIAQPRAGASTAVP 542
Db 487 KQAAWIGAYEOWNVDIGLEGLSGHAQIGKGMWAMPDMWAAMLSEKLAHPKAGANTAVP 546
Qy 543 SPTAATLHALYHVDVAAGVQGLAGKRATIEOLLTIPLAKELAWAPDEIRREVDNNQ 602
Db 547 SPTAATLHATHYKVDVAAGVQGLAGKRATIEOLLTIPLAKELAWAPDEIRREVDNNQ 606
Qy 603 SILGVYVRWDQGVCSKVPDIHVALMEDRATIRISQLLANMLRHQVITSADVRLASL 662
Db 607 GILGVYVRWDQGVCSKVPDINNIGLMEDRATIRISAQHM 666
Qy 663 RMAPLVRQNDGVAIRPMAPNFDSDIAFLAAQELISGAQOPNGYTPILHRRRRREFK 721
Db 667 RMAAIVDQKNEGDPYRPMADFDKSIQAFQACDLVFKGREQPNGYTEPVLHRRRLLELK 725

RESULT 7

I40715
malate synthase (EC 4.1.3.2) - Corynebacterium glutamicum
C:Species: Corynebacterium glutamicum
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 05-May-2000
C:Accession: I40715; I40836
R:Reinisch, D.J.; Ekman, B.J.; Sahm, H.
Microbiology 140, 3099-3108, 1994
A:Title: Malate synthase from Corynebacterium glutamicum: sequence analysis of the gene
A:Reference number: I40715; MUID:951111631; PMID:7812449
A:Accession: I40715
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-739 <RES>
A:Cross-references: EMBL:X78491; NID:9530011; PIDN:CAA55243.1; PID:9530012
R:Lee, H.
J. Microbiol. Biotechnol. 4, 256-263, 1994
A:Title: Molecular characterization of aceB, a gene encoding malate synthase in Coryneba
A:Reference number: I40836
A:Accession: I40836
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-739 <RES>
A:Cross-references: GB:L27123; NID:9653573; PIDN:AAA68074.1; PID:9853574

A:Reference number: S51788; MUID:95010032; PMID:7925370

A:Accession: S51788

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-723 <MOU>

A:Cross-references: EMBL:X74547; NID:9517246; PIDN:CAA52639.1; PID:9517247

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shaoh, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F65083

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-723 <BLAT>

A:Cross-references: GB:AE000380; GB:U00096; NID:91789344; PIDN:AAC76012.1; PID:91789348;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: glcB

C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 56.5%; Score 2154.5; DB 2; Length 723;

Best Local Similarity 58.0%; Pred. No. 9.7e-136;

Matches 419; Conservative 115; Mismatches 174; Indels 15; Gaps 5;

QY 4 VSVGNLRIRVLYDFVNEALPGTDIDPDSFAGVGVADLPONCALNARDELQAQI 63

DB 5 ITQRLRIDANFKFVDEVLPGTGLDAAAFWRNFDIVHDLAPENQLLAERDRQAAL 64

QY 64 DKWHRRIEVIDMAYRQFTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVWPVNA 123

DB 65 DEWHRSPGPVKDAAKYSFLRELGLVPOPERVTVETGIDSEITTSQAGPQLVWPVNA 124

QY 124 RFLANANARGLSYDALYGTVDVTPETDGAEGKTYNKVRGDKVIARFELDDSVPLSS 193

DB 125 RYALNANARGLSYDALYGSIIIPQEGAMVSG--YDPQGEQVIAWRFLDESUPLEN 192

QY 184 GSFQDGTFTVQDQQLVVALPD-KSTGLANPGQFAGYTGAAESPTSVLLINHGHIETILI 242

DB 183 GSYQDVAFKVVQKLRQLKNGKETTLLTPAQVGVGDAAPTCLLKNNGHIELQI 242

QY 243 DPESQVTTDRAGVKDVLSSAITTIMDFEDSVAAADKVLGYRNWGLNKGDLAAAV 302

DB 243 DANGRIGKDDPAHINDIVIAAIGTILDCEDSVAAVDAEDKILLRYNLLGLMQGTLQEKM 302

QY 303 DKDGTAFRLVLRDRNTYAPCGGQFTLPGRLSMFVRNVGHMTNDATVDTDGSEVFEGIM 362

DB 303 EKNRQIVKLRNDRPHVTAADGSIHLGRSLFPIRVNGHMLTPIVNDSEGNIEPGL 362

QY 363 DALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKHGAFAVFTCELSRVEDVLGLP 422

DB 363 DGVMTGAIALYDLKVKQ-----NSRTGSIYIVKPKHGPQEVAFANKLFTRIETMLGMA 416

QY 423 QNTWKIGTMDERITTVNLKACIAAADRVVFIINTGFLDRTGDEIHTSMEAGPMVRKGTM 482

DB 417 PNTLKGIMLDERITSLNRSCTIAQRNVAFINTGFLDRTGDEMHSWAGPMVRKQW 476

QY 483 KSQPMILAYEDPHNVDAAGLAFSGRAQVGKGMVTMTLMADWVETKIAQPRAGASTAWVP 542

DB 477 KSTPMKAYERNVLSGLFCGLRGKAQIGKGMAMPDMADMYSQKQDQLRAGANTAWVP 536

QY 543 SPTAATLHALHYQDVAAVQOGLA-----GKRATTIEQLLTPLAKELAWADETREEV 597

DB 537 SPTAATLHALHYQTNVQSVQNTAQTEFNAEFEPFLDLDLTIPVAENANWSAQEQSEL 596

QY 598 DNNQCSILGYVVRVVDQGVGSKVPDIHDVVALMEDRATLRISSQLLANLWRHGVITTSADY 657

DB 597 DNNVQILGYVVRVVEQIGGSKVPDIHNVALMEDRATLRISSQHIANLWRHGLITKEQV 656

QY 658 RASLERMAPLVDQRAGDVAYRMAPNPFDDSDIAFLAQELILSGAQPNGYTPEILHR-R 716

DB 657 QASLENMAKVVDQQRAGDYPYRMAPNAGNFANSACFAKASDLIFLGVKQPNGYTEPLLHAWR 716

QY 717 RRE 719

DB 717 LFE 719

RESULT 10

S17774

malate synthase (EC 4.1.3.2) - *Neurospora crassa*

C:Species: *Neurospora crassa*

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-May-2000

C:Accession: S17774

R:Sandeman, R.A.; Hynes, M.J.; Fincham, J.R.S.; Connerton, I.F.

Mol. Gen. Genet. 228, 445-452, 1991

A:Title: Molecular organisation of the malate synthase genes of *Aspergillus nidulans* and

A:Reference number: S17773; MUID:91375430; PMID:1832736

A:Accession: S17774

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-542 <SAN>

A:Cross-references: EMBL:X56672; NID:92974; PIDN:CAA39994.1; PID:92975

C:Genetics: 447/3

A:introns: 447/3

C:Superfamily: malate synthase

C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 5.4%; Score 205.5; DB 2; Length 542;

Best Local Similarity 24.4%; Pred. No. 6.6e-06;

Matches 135; Conservative 64; Mismatches 205; Indels 149; Gaps 33;

QY 194 VODGOLVVALPKSTGLANPGQFAGYTGAAESPTSVLLINHGHIETILDPESQVGTDR 253

DB 58 IDRGVLDPDFLPETKHIRENP----TWKGAA--PAAPPLVDRRVEV-----TGPTDR 102

QY 254 AGVKDVLIESAITTIMDFEDSVAAVDAADKVLGYRNWGLNKGD-----LAAAVD-KDG 306

DB 103 KVVNALNSDVVTYNADFEDSSAP-----TWANMVNGQNLVDATRQIDFKQG 151

QY 307 TAFRLVLRDRNTYAPCGGQFTLPGRLSMFVRNVG-HLMTNDATVDTDGSEVFEGIMDAL 365

DB 152 PREYK-LATDR-----TLP--TLIVRPGWHL--EEKVTTIDGEPVSGSLFD-- 193

QY 366 FTGLIAHGLKASDVNGPLINSRTGSIYIVKPKHGAFAVFTCELSRVEDVLGLPONT 425

DB 194 -FGLYFFFNK-----ELVQRGFGP-YFYPPKMHSEHLEARLWDAFNLAQDYVGIPLST 245

QY 426 MKIGIMDEERTTVNLKACIKAAADRVVFIINTGFLDRTG-DEIHTSMEAGPMV--- 477

DB 246 I-----RGTVLEITAAFEDEIIFELRNHTSGLNRGGWDYIF-----PFIKEV 290

QY 478 -----RKGTWKSQFWILAYED-----HNV-----DAGLAAGFSGRAQVGKGMVTMT 518

DB 291 RFPNFVLPRSDVTMTVPFMEAYVKLLIKTLHLVWHAAGCAAQIPIKDDKAAANDKAME 350

QY 519 ELMAWMTETKIAQPRAGASTAWVSPSTAATL-----HALHYQDVAAVQOGL 566

DB 351 GVRAD-----KIREAPAGHDGTWVAHPALASIALEVFNKHMTPTNQLFNRREDVKIGQDL 406

QY 567 AGKRATTIEQLLTPLAKELAWADETREEVNDCQSILGYVVRVVDQGVGSKVPDIHD 626

DB 407 LN-----MNVPGSS-----TEDGIRK-----NLNTGLGYTEPMI-RGVGC--VPKHP 446

QY 627 VALMEDRATLRISSQLLANLWRHGVITTS-----ADVRASLERMAPLVDQRAGDVAYRMAP 682

DB 447 ---QEDALTAEVSRQLWQWVHRVTTAEGKHVYKRYPL-KLLKEADRQR---LAKAPQG 499

QY 683 PNFDDSIAPLAAQ 695

DB 500 NKENLAAQYFASQ 512

RESULT 11

D91245

malate synthase A [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD 0509952)

C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D91245
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gawawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hatcori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D91245
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-533 <HAY>
A:Cross-references: GB:IA000007; PIDN:BA38355.1; PID:gl3364408; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs4932
C:Superfamily: malate synthase

Query Match 5.0%; Score 190.5; DB 2; Length 533;
Best Local Similarity 22.8%; Pred. No. 6.4e-05;
Matches 117; Conservative 67; Mismatches 241; Indels 89; Gaps 21;

QY 159 YNKVRGDKVIYARKFLDDSVPLSSGSGFGDATGFTVQDQOLVVALPKDSTGLANPGQFAG 218
DB 17 YGEOEKQILTAZAVFELTV-----THFTPQRNKLAAARIQQQODIDN-GTLPD 65

QY 219 YTGAAESPTSVLLINHLHIEILIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAAV 278
DB 66 FSETASIRDADWKIRGIPADLEDRVEITGSEVERKWINALNANVKVFMADFEEDSLAP- 124

QY 279 DAADKVLGYRNWGLNKGDLAAVADKGTAFIL-VLNDRNRYTAPGGQFTLPGRSLMFV 337
DB 125 -----DW---NK-----VIDGQINLRDAVNGTISYTNAGKIYQKPNPAVLI 164

QY 279 DAADKVLGYRNWGLNKGDLAAVADKGTAFIL-VLNDRNRYTAPGGQFTLPGRSLMFV 337
DB 125 -----DW---NK-----VIDGQINLRDAVNGTISYTNAGKIYQKPNPAVLI 164

QY 338 RNUGHLMNTDAIVDTDGSEVFEGIMDALFT-GLIAIHGLKA--SDVNGPLINSRTGSIYI 394
DB 165 CRVGLHLPEKHVTWRG-----EALPGSLFDALYFFHNYQALLAKGSGP-----YF 211

QY 395 VKPQHGPAEVAFTCELFSEVEDVLGPONTWKIGIMDEERTTVNLKACIKAAADRVWF 454
DB 212 YLPKTSQWQEAAMWSEVFSYAEDRNLPRGTIKATLIELTLPVAFQMDIHALRDHVG 271

QY 455 INTGFLDRTGDEHTSMEAGPMV---RKGTMSQPMILAYED-----HNVDA-----GLA 501
DB 272 LNCGRWDYIFSYIKTLKNYPDRVLPDRQAVTMDKPFNLNAYSRLLIKTCHRGAFAMGMA 331

QY 502 AGFSGRAQVKGKMTWELMADVETKIAQPRAGASTAWVPSPTAATLHALHYHQVDVA- 560
DB 332 AFIPSKDEEHNQ-VLNKVKAD---KSLAANNHGDGTWIAHPGLA-----DTAM 376

QY 561 AVQOGLAGKRATIEQL--LTIPLAKELAWP---DEIREEDNNCSILGYVVRWVDQ 615
DB 377 AVFNDILGSRKNQLEVRREQDAPITADQLLAPCDGERTEGMRANIRVAVQVIEAWI-SG 435

QY 616 VGCSKVPDIHDVALMEDRATLRISQLLANWLH 649
DB 436 NGCVPI-----YGLMEDAATAEISRTSIWQWIH 464

RESULT 12
B86093
malate synthase A [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B86093
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B86093
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-533 <STO>
A:Cross-references: GB:AE005174; NID:gl2518944; PIDN:AAGS9206.1; GSPDB:GN00145; UWGP:Z56
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: aceB
C:Superfamily: malate synthase

Query Match 5.0%; Score 190.5; DB 2; Length 533;
Best Local Similarity 22.8%; Pred. No. 6.4e-05;
Matches 117; Conservative 67; Mismatches 241; Indels 89; Gaps 21;

QY 159 YNKVRGDKVIYARKFLDDSVPLSSGSGFGDATGFTVQDQOLVVALPKDSTGLANPGQFAG 218
DB 17 YGEOEKQILTAZAVFELTV-----THFTPQRNKLAAARIQQQODIDN-GTLPD 65

QY 219 YTGAAESPTSVLLINHLHIEILIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAAV 278
DB 66 FSETASIRDADWKIRGIPADLEDRVEITGSEVERKWINALNANVKVFMADFEEDSLAP- 124

QY 279 DAADKVLGYRNWGLNKGDLAAVADKGTAFIL-VLNDRNRYTAPGGQFTLPGRSLMFV 337
DB 125 -----DW---NK-----VIDGQINLRDAVNGTISYTNAGKIYQKPNPAVLI 164

QY 338 RNUGHLMNTDAIVDTDGSEVFEGIMDALFT-GLIAIHGLKA--SDVNGPLINSRTGSIYI 394
DB 165 CRVGLHLPEKHVTWRG-----EALPGSLFDALYFFHNYQALLAKGSGP-----YF 211

QY 395 VKPQHGPAEVAFTCELFSEVEDVLGPONTWKIGIMDEERTTVNLKACIKAAADRVWF 454
DB 212 YLPKTSQWQEAAMWSEVFSYAEDRNLPRGTIKATLIELTLPVAFQMDIHALRDHVG 271

QY 455 INTGFLDRTGDEHTSMEAGPMV---RKGTMSQPMILAYED-----HNVDA-----GLA 501
DB 272 LNCGRWDYIFSYIKTLKNYPDRVLPDRQAVTMDKPFNLNAYSRLLIKTCHRGAFAMGMA 331

QY 502 AGFSGRAQVKGKMTWELMADVETKIAQPRAGASTAWVPSPTAATLHALHYHQVDVA- 560
DB 332 AFIPSKDEEHNQ-VLNKVKAD---KSLAANNHGDGTWIAHPGLA-----DTAM 376

QY 561 AVQOGLAGKRATIEQL--LTIPLAKELAWP---DEIREEDNNCSILGYVVRWVDQ 615
DB 377 AVFNDILGSRKNQLEVRREQDAPITADQLLAPCDGERTEGMRANIRVAVQVIEAWI-SG 435

QY 616 VGCSKVPDIHDVALMEDRATLRISQLLANWLH 649
DB 436 NGCVPI-----YGLMEDAATAEISRTSIWQWIH 464

RESULT 13
AD1011
malate synthase (EC 4.1.3.2) - Salmonella enterica subsp. enterica serovar Typhi (strain
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AD1011
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AD1011
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-533 <PAB>
A:Cross-references: GB:AL513382; PIDN:CAD09189.1; PID:gl6505193; GSPDB:GN00176
C:Genetics:
A:Gene: STY4401
C:Superfamily: malate synthase
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 5.0%; Score 190; DB 2; Length 533;

Search completed: November 21, 2003, 16:09:50
Job time : 18.4376 secs

This Page Blank (uspio)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 8.72562 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-2

Perfect score: 3810

Sequence: 1 TDRSVGNLRIARVLYDFVN.....KARAAEKAPSPRAGDDAAR 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	100.0	741	1 MASZ_MYCTU	Q50596 mycobacteri
2	3117.5	81.8	731	1 MASZ_MYCLE	Q32913 mycobacteri
3	2634	69.1	724	1 MASZ_RHOFA	Q9a655 rhodococcus
4	2569	67.4	725	1 MASZ_PSEAE	Q9a636 pseudomonas
5	2541	66.7	725	1 MASZ_PSEPK	Q860x8 pseudomonas
6	2497	65.5	725	1 MASZ_PSEFL	Q05137 pseudomonas
7	2489	65.3	725	1 MASZ_PSESM	Q88ab2 pseudomonas
8	2416	63.4	721	1 MASZ_RHILO	Q98dk4 rhizobium l
9	2372.5	62.3	731	1 MASZ_AGR5	Q8u185 agrobacteri
10	2367.5	62.1	723	1 MASZ_RHILV	Q937w7 rhizobium l
11	2351.5	61.7	723	1 MASZ_RHIME	Q92ta4 rhizobium m
12	2318	60.8	728	1 MASZ_BRUME	Q8yir3 brucella ne
13	2313	60.7	728	1 MASZ_BRUSU	Q8fz50 brucella su
14	2289	60.1	738	1 MASZ_CORGL	P42450 corynebacte
15	2273	59.7	748	1 MASZ_COREF	Q8fnb3 corynebacte
16	2272.5	59.6	721	1 MASZ_BRAJA	Q89ue3 bradyrhizob
17	2256.5	59.2	727	1 MASZ_BACHD	Q9kb03 bacillus ha
18	2215	58.1	725	1 MASZ_PSESM	Q87272 pseudomonas
19	2168.5	56.9	722	1 MASZ_ECOL6	Q8fdn6 escherichia
20	2156.5	56.6	722	1 MASZ_SHIFL	P59663 shigella fl
21	2154.5	56.5	722	1 MASZ_ECOLI	P73330 escherichia
22	205.5	5.4	542	1 MASZ_NEUCR	P28345 neurospora
23	193.5	5.1	1005	1 GCP_CABEL	Q10663 caenorhabdi
24	189.5	5.0	533	1 MASZ_ECOLI	P08997 escherichia
25	185	4.9	551	1 MASZ_CANTR	Q02116 candida tro
26	184.5	4.8	541	1 MASZ_MYXXA	P95329 myxococcus
27	175	4.6	528	1 MASZ_EMENI	P28344 emericella
28	171.5	4.5	559	1 MASZ_MAIZE	P49081 zea mays (m
29	171	4.5	554	1 MASZ_YEAST	P21826 saccharomyc
30	165	4.3	554	1 MASZ_YEAST	P30952 saccharomyc
31	160	4.2	555	1 MASZ_PICAN	P21360 pichia angu
32	153.5	4.0	541	1 MASZ_STRCL	Q92477 streptomyce
33	151.5	4.0	564	1 MASZ_SOYBN	P45458 glycine max

34 150.5 4.0 561 1 MASZ_BRANA
35 149.5 3.9 566 1 MASZ_RAPSA
36 149 3.9 566 1 MASZ_CUCWA
37 145 3.8 543 1 MASZ_STRAE
38 144.5 3.8 568 1 MASZ_CUGSA
39 138 3.6 567 1 MASZ_RICCO
40 128.5 3.4 597 1 MEHL_ECOLI
41 128.5 3.4 808 1 SEAD_MYCTU
42 125.5 3.3 1039 1 AG43_ECOLI
43 124.5 3.3 827 1 CSG_HALVO
44 123.5 3.2 914 1 IF2_CHLTE
45 121 3.2 567 1 MASZ_GOSHI

ALIGNMENTS

RESULT 1
MASZ_MYCTU
ID MASZ_MYCTU STANDARD; PRT; 741 AA.
AC Q50596;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase C (EC 2.3.3.9).
GN GLCB OR RV1837C OR M1885 OR M1C1A11.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Ailand D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
Decher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bisai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. G1CB subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see [http://www.ebi.ac.uk/ebis-sib.ch/](http://www.ebi.ac.uk/ebis-sib.ch/announcement/)
or send an email to license@ebi-sib.ch).
CC EMBL; 278020; CAB01465.1; -.
CC EMBL; AE007047; AAK46156.1; -.

```
DR PIR; F70722; F70722.
DR PDB; IN81; 18-DEC-02.
DR PDB; IN8W; 18-DEC-02.
DR TIGR; MT1885; -.
DR Tuberculin; Rv1837c; -.
DR HAMAP; MF_00641; -.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn G; 1.
KW Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome; 3D-structure
FT ACT_SITE 339 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 633 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 741 AA; 80403 MW; A32F54E0FE8B7C64 CRC64;

Query Match 100.0%; Score 3810; DB 1; Length 741;
Best Local Similarity 100.0%; Pred. No. 2.5e-241;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDRVSGNLRIRARVLYDFVNNALPGTIDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 60
Db 2 TDRVSGNLRIRARVLYDFVNNALPGTIDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 61
QY 61 A QIDKWHRRVIEPIEDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTAGPOLVVPV 120
Db 62 A QIDKWHRRVIEPIEDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTAGPOLVVPV 121
QY 121 LNARFALNAANARWGSGLYDALYGTDIVIPETDGAEGKGYTNKVRGDKVIAYARKFLDDSV 180
Db 122 LNARFALNAANARWGSGLYDALYGTDIVIPETDGAEGKGYTNKVRGDKVIAYARKFLDDSV 181
QY 181 LSSSFGDGTFTVQDQGLVALPKSTGLANPQFAGYGAESPVSILLINHLHIEI 240
Db 182 LSSSFGDGTFTVQDQGLVALPKSTGLANPQFAGYGAESPVSILLINHLHIEI 241
QY 241 LIDPESQVGTTRDAGVXDVIIESAITTIMPDSVAADKVLGVNMLGLNKGDLLA 300
Db 242 LIDPESQVGTTRDAGVXDVIIESAITTIMPDSVAADKVLGVNMLGLNKGDLLA 301
QY 301 AVDXDGTAFRLNDRNYTAPGGGQFTLPGRSLNFMVNRVGHLMNDIAIVDTGSEVPEG 360
Db 302 AVDXDGTAFRLNDRNYTAPGGGQFTLPGRSLNFMVNRVGHLMNDIAIVDTGSEVPEG 361
QY 361 IMDALFTGLIAIHGLKASDVNGPLINSGTSGIYIVKPMGPAEVAFTCELFSRVEDVLG 420
Db 362 IMDALFTGLIAIHGLKASDVNGPLINSGTSGIYIVKPMGPAEVAFTCELFSRVEDVLG 421
QY 421 LPQNTMKIGIMDEERTTVNLKACIKAAADRVVFINTGFLDRGTDEIHTSMEAGPMVRKG 480
Db 422 LPQNTMKIGIMDEERTTVNLKACIKAAADRVVFINTGFLDRGTDEIHTSMEAGPMVRKG 481
QY 481 TMKSQPTILAYEDHNVDAGLAAGSGRAQVCKGMWMTLMADWVETKIAQPRAGASTAW 540
Db 482 TMKSQPTILAYEDHNVDAGLAAGSGRAQVCKGMWMTLMADWVETKIAQPRAGASTAW 541
QY 541 VPSPTAATLHALHYQVDVAQVQGLAGKRATIEQLTIPLAKELAWAPDEIREEDNN 600
Db 542 VPSPTAATLHALHYQVDVAQVQGLAGKRATIEQLTIPLAKELAWAPDEIREEDNN 601
QY 601 COSILGVVVRWDQVCGSKVPDTHDVALMEDRATLRISQLLANWLHGVITSADVRAS 660
Db 602 COSILGVVVRWDQVCGSKVPDTHDVALMEDRATLRISQLLANWLHGVITSADVRAS 661
QY 661 LERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAQAEILILSGAQPNGCYTEPIILHRRREF 720
Db 662 LERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAQAEILILSGAQPNGCYTEPIILHRRREF 721
QY 721 KARAAEKPAFSDRAGDDAAR 740
Db 722 KARAAEKPAFSDRAGDDAAR 741

RESULT 2
MASZ MYCLE
ID MASZ MYCLE STANDARD; PRT; 731 AA.
AC O329T3.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR ML2069 OR MLCB1788.27.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxid=1789;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity). GlcB subfamily.
CC -!- SIMILARITY: Belongs to the malate synthase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL008609; CAAL5459.1; -.
DR EMBL; AL583924; CAC31024.1; -.
DR FIR; T44752; T44752.
DR HSSP; P37330; 1D8C.
DR Leptoma; ML2069; -.
DR HAMAP; MF_00641; -.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn G; 1.
KW Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 340 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 638 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 731 AA; 80141 MW; 3878CADA5DB416C CRC64;

Query Match 81.8%; Score 3117.5; DB 1; Length 731;
Best Local Similarity 82.0%; Pred. No. 4.3e-196;
Matches 597; Conservative 54; Mismatches 7; Indels 5; Gaps 2;

QY 1 TDRVSGNLRIRARVLYDFVNNALPGTIDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 60
Db 2 TDRVSGNLRIRARVLYDFVNNALPGTIDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 61
QY 61 A QIDKWHRRVIEPIEDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTAGPOLVVPV 120
Db 62 A QIDKWHRRVIEPIEDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTAGPOLVVPV 121
QY 121 LNARFALNAANARWGSGLYDALYGTDIVIPETDGAEGKGYTNKVRGDKVIAYARKFLDDSV 180
```

Db 122 LNARFALNAANARWGSYLDALYGTDTIPETEGAEGKSEYNKIRGDKVIAVARKFMDQAVP 181
 QY 181 LSSGSGFDATGFTVQDQOLVVALPDKSTGLANPQAGTGAABST-SVLLNHGLHIE 239
 Db 182 LADSWTNAGVSIFDQQLQIATGNTSTGLASKEKFNVRNQRSSNWSVLLANHGLHIE 241
 QY 240 ILIDPESQVQTTDRAGVKDVIKESAITTTIMDFDSVAADDAADKVLGYRNWGLNKGDLA 299
 Db 242 VLIDPESPIKTPDVGKIVKILESAITTTIMDFDSVTAVDAADKVGYNWGLNKGDLT 301
 QY 300 AAYDKDGTAFRLVNRDRNTATPGGQFTPLGRSLFVRNVGHLMNDIAIV---DTDGS 355
 Db 302 ERYNKDKTFTRVINADRSYTTDPGGELTLPGRSLFVRNVGHLMNDIAIV---DTDGS 361
 QY 356 EVEGIMDLFTGLIAIHGLKASDVNGPLNSITGSIYIVKPMHGPAAEVAFTCELFSSV 415
 Db 362 EVEGIIIDAVFTGLAAIHGLKGTGANGPLNSITGSIYIVKPMHGPAAEVAFTCELFSSV 421
 QY 416 EDVLGILPQNTWKIGIMDEERTTNLKACIKAAADRVVFNFTGFLDRTGDEIHTSMEAGP 475
 Db 422 EDVLGILPQNTWKIGIMDEERTTNLKACIKAAADRVVFNFTGFLDRTGDEIHTSMEAGP 481
 QY 476 MVKSGTKMSQPTLAVEDHNVDAGLAAGFSGRAQVKGKMTWTELMADWVETKIAQPRAG 535
 Db 482 MIRKGAKNSTWIKAYEDANVDIGLAAGFKGAQIGKGMWMTLMADWVEQKIGQPRAG 541
 QY 536 ASTAWPSPATATLHALHYQVDVAAVQOGLAGKREATTIEQLITPLAKELAWAPDEIRE 595
 Db 542 ATTAWPSPATATLHAMHYQVDVAAVQOGLTQRRATVDQLITPLAKELAWAPDEIRE 601
 QY 596 EVDNQCQSLIGYVVRVWDQVGSQVDPDHDVALMEDRATLRISSOLLANWLHGVITSA 655
 Db 602 EVDNQCQSLIGYVVRVWDQVGSQVDPDHDVALMEDRATLRISSOLLANWLHGVITSE 661
 QY 656 DVSASLERMAPLVDQVQNAEDPARVPAPNFDDSIAPLAQELILSGAQQPNGYTEPILHR 715
 Db 662 DVSASLERMAPLVDQVQNAEDPARVPAPNFDDSIAPLAQELILSGAQQPNGYTEPILHR 721
 QY 716 RRREFKAR 723
 Db 722 RRREFKAR 729

RESULT 3
 MASZ_RHOFA STANDARD; PRT; 724 AA.
 AC Q9AES5;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Malate synthase G (EC 2.3.3.9).
 GN GLCB OR VICA.
 OS Rhodococcus fascians.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Nocardiaceae; Rhodococcus.
 OX NCBI_TaxID=1828;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D189;
 RA Verecke D.M., Cornelis K., Van Montagu M., El Jaziri M., Hoisters M.,
 RA Goethals K.;
 RA "Characterization of a chromosomal locus that affects pathogenicity in
 RT Rhodococcus fascians";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
 CoA.
 CC -1- PATHWAY: Glyoxylate bypass; second step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ301559; CAC35701.1; -
 DR HSSP; P37330; 1D8C
 DR HAMAP; MF_00641; 1
 DR InterPro; IPR001465; Malate_synthase.
 DR InterPro; IPR006253; Malate_synthase.
 DR Pfam; PF01274; Malate_synthase; 1.
 DR TIGRFAMs; TIGR01345; malate syn G; 1.
 KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
 FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
 SQ SEQUENCE 724 AA; 78609 MW; F889FE883890995E CRC64;
 Query Match 69.1%; Score 2634; DB 1; Length 724;
 Best Local Similarity 69.0%; Pred. No. 1.6e-164;
 Matches 499; Conservative 82; Mismatches 138; Indels 4; Gaps 2;
 QY 1 TDRSVGNLRRLARVLVDFVNNALPGTIDDPDSWAGVDKVVADLTTPONQALLNARBLO 60
 Db 2 TDRVQAGLQVAKVLFDFVEKEALPGTIDDLSEAFWAGAAASVIADLAPKXKALLAVRDEIQ 61
 QY 61 AQIDKWHRRRVEIPIDMDAYROFLTETIGYLLPEPDDFTITTSVGDVDAEITTTAGPOLVVPV 120
 Db 62 GKVDAMHGEHAGAEVDRAAYKAFLEIGYLLDEADFOIHTSGVDTEITTTAGPOLVVPV 121
 QY 121 LNARFALNAANARWGSYLDALYGTDTIPETDGAEGKPTYNKRGDKVIAYARKFLDSDVP 180
 Db 122 LNARFALNAANARWGSYLDALYGTDAIPETDGAEGKPTYNKRGDKVIAYARKFLDSDVP 181
 QY 181 LSSSGFGDATGFTVQDQOLVVALPDKST-GLANPQAGTGAABST-SVLLNHGLHIE 239
 Db 182 LSSSHVGTGTVVDAASLTVTLAGDSTVGLKDSQLLGYQGTDPATLFFVHGLHFE 241
 QY 240 ILIDPESQVQTTDRAGVKDVIKESAITTTIMDFDSVAADDAADKVLGYRNWGLNKGDLA 299
 Db 242 ILIDPESPIKTPDVGKIVKILESAITTTIMDFDSVAADDAADKVLGYRNWGLNKGDLT 301
 QY 300 AAYDKDGTAFRLVNRDRNTATPGGQFTPLGRSLFVRNVGHLMNDIAIVDTDGSEVFE 359
 Db 302 BEVSKGGKTFTRAMNKORTYTSVDGSELTLHGRSLLFVRNVGHLMNDIAIVDADGNEVPE 361
 QY 360 GIMDALFTGLIAIHGLKASDVNGPLNSITGSIYIVKPMHGPAAEVAFTCELFSSVDEVL 419
 Db 362 GILDALFTSLAGLHSLTDPDNV--LSNSRTGSLYIVKPMHGPDEVAFTALFGRVEQVL 418
 QY 420 GLPQNTMKIGIMDEERTTNLKACIKAAADRVVFNFTGFLDRTGDEIHTSMEAGPMVRK 479
 Db 419 GLPTNTLVKIGIMDEERTTNLKACIQAASRVVFNFTGFLDRTGDEIHTSMEAGPMVRK 478
 QY 480 GTMKSQPIALAYEDHNVDAGLAAGFSGRAQVKGKMTWTELMADWVETKIAQPRAGASTA 539
 Db 479 GAKKGEKWIAYEDFNVDVTGLAAGLQKGAQIGKGMWAMPDLNMDHMLEQKIGHPKAGANTA 538
 QY 540 WVPGSTAATLHALHYQVDVAAVQOGLAGKREATTIEQLITPLAKELAWAPDEIREEDVN 599
 Db 539 WVPGSTAATLHALHYKVDVFAHQETAKAKRATVDEILEIPLAPSTDWTDEEKQNELDN 598
 QY 600 NCQSTLGVVVRVWDQVGSQVDPDHDVALMEDRATLRISSOLLANWLHGVITSAADVA 659
 Db 599 NSQSTLGVVVRVWDQVGSQVDPDHDVALMEDRATLRISSOLLANWLHGVITSAADVA 658
 QY 660 SLERMAPLVDQVQNAEDPARVPAPNFDDSIAPLAQELILSGAQQPNGYTEPILHRRRE 719
 Db 659 SLKRMAPVVDQVQNAEDPYKELAFDFTDNTIAFQAASDLIFQGTSPQPNGYTEPILHRRRE 718
 QY 720 FKA 722
 Db 719 YKA 721

242 IQIDASTPVGQTDAGVKVQVMEALTTIMDCEDSVAADDDKVVYIRNWLGLMKGDLA 301
 300 AAVDKDGTAFRLVLRNDRNYTAPGGQFTLPGSLMEFVRNVGHLMTNDAIVDTGGSEVFE 359
 302 EEVAKGGKTFRTMNPDRVYTVGQDVTLHGRSLLFVRNVGHLMTIDAILDKAGNEVPE 361
 360 GIMDALFTGLIAIHLKASDVNGPLINSRTGSIYIVKPKMGHPAEVAFTCELSFVEDVL 419
 362 GILDGLTSLAAIHSLNG---NSRKNRSRTGSIYIVKPKMGHPAEVAFTCELSFVEDVL 418
 420 GLPONTMKGIMDEERRTTNKLKACIKAAADRVVFIINTGFLDRTGDEIHTSMGAGPMVRK 479
 419 NLPRNTLKVGMIDERRTTNKLKACIKAAASERVFIINTGFLDRTGDEIHTSMGAGPMVRK 478
 480 GTMKSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADWVETKIAOPRAGASTA 539
 479 AAMTEKWIYAYENWVNDIGLSTGLQGRAQIGKGMWAMPDLMAAAMLEOKIAHPLAGANTA 538
 540 WVPSPATAATLHALVHVQDVAAVOGLAGKERRATIEOLLTIPLAKELAWAPDEIRREEDVN 599
 539 WVPSPATAALHALVHVQDVFAQAELAKRERASVDDILTIPLAKNTDMSSEELRNELDN 598
 600 NCOSILGVVVRWVGCGSKVPDIHDVVALMEDRATLRISQILLANLWLRHGVTSDVRA 659
 599 NAQOILGVVVRWIDQGVGSKVPDINDVGLMEDRATLRISQILLANLWLRHGVTSDQVME 658
 660 SLERMAPLVDRQAGDVAYRPMNFDSDIAFLAQELILSGAOPNGYTEPIHLRRRRE 719
 659 SLKRMAPVVDQRQAGDALLYRPLAPDFDSNIAFQAAVELVIEGTQKPNGYTEPVLHRRRE 718
 720 FKAR 723
 719 FKAR 722
 RESULT 7
 MAZ1 PSESM STANDARD; PRT; 725 AA.
 ID MAZ1 PSESM
 AC Q88A2;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Malate synthase G 1 (EC 2.3.3.9).
 GN GLCB1 OR GLCB-1 OR PSPT00480.
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
 RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
 RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
 RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
 RA White O., Fraser C., Collier A.;
 RA "Complete sequence of Pseudomonas syringae".
 RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 RL
 CC -! CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
 CC CoA.
 CC -! PATHWAY: Glyoxylate bypass; second step.
 CC -! SUBUNIT: Monomer (By similarity).
 CC -! SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -! SIMILARITY: Belongs to the malate synthase family. GicB subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

EMBL: AE016857; AAC54024.1; --
 TIGR: PSPT00480; --
 DR HAMAP; MF_00641; --; 1.
 KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
 KW Complete proteome.
 FT ACT SITE 340 CATALYTIC BASE (BY SIMILARITY).
 FT ACT SITE 631 CATALYTIC ACID (BY SIMILARITY).
 FT ACT SITE 631 CATALYTIC ACID (BY SIMILARITY).
 SQ SEQUENCE 725 AA; 79143 MW; F1993364E8083660 CRC64;
 Query Match 65.3%; Score 2489; DB 1; Length 725;
 Best Local Similarity 65.3%; Pred. No. 4.8e-15;
 Matches 473; Conservative 97; Mismatches 150; Indels 4; Gaps 2;
 QY 1 TDVSVGNLRVLRVLYDFVNEALPGTIDIDPSWAGVDKVVADITPONQALLNARDELQ 60
 DB 2 TEVVOVGDLQVAVLDFVQNEATEFTGVDAGAFWAGADQLIHLAPKNKALLAQRDDELQ 61
 QY 61 AOIDKWHRRRVTEPIDMAYRQFLTEIGYLLPEPDDFTTITSGVDABITTTAGPOLVVPV 120
 DB 62 AOIDAHOSRAGQAHDASAYKAFLOEIGYLLPEAADFOITTCNVDEEATWAGPOLVVPV 121
 QY 121 LNARFALNAANARWGLSYDALYGTDTVIPETDCAEKGPTYNKVRGDKVIAYARKFLDDSV 180
 DB 122 MNARFALNAANARWGLSYDALYGTDTVISEGGAEGKGNKVRGDKVIAYARAFLDQAP 181
 QY 181 LSSGSGDGTFTVQDQGLVWALPKS-TGLANPCQFAGYTGAAESPTSVLLINHLHIE 239
 DB 182 LAAGSHVDSTAYKLIDGRVLISLKGGSNTGLRDDAQLVGFQGDASAPFAVLFKHGLHFE 241
 QY 240 ILIDPESQVTTDRAGVKDVILESATITIMDPEDSVAADKVLGVRNVLGMLKGDIA 299
 DB 242 LQIDRAASPVGQTDPAQVDIVMEALTTIMDCEDSIAAVDADKKVVVRNVLGMLKGDIV 301
 QY 300 AAVDKDGTAFRLVLRNDRNYTAPGGQFTLPGSLMEFVRNVGHLMTNDAIVDTGGSEVFE 359
 DB 302 ESVKSGKETFRTMNPDRVYTVGQDVTLHGRSLLFVRNVGHLMTIDAILDKAGNEVPE 361
 QY 360 GIMDALFTGLIAIHLKASDVNGPLINSRTGSIYIVKPKMGHPAEVAFTCELSFVEDVL 419
 DB 362 GILDGLTSLAAIHSLNGNTRS---NSRSGSMYIVKPKMGHPQEAFTNEUFGRIEQL 418
 QY 420 GLPONTMKGIMDEERRTTNKLKACIKAAADRVVFIINTGFLDRTGDEIHTSMGAGPMVRK 479
 DB 419 GLPNTLKVGMIDERRTTNKLKACIQAAASERVFIINTGFLDRTGDEIHTSMGAGPMVRK 478
 QY 480 GTMKSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADWVETKIAOPRAGASTA 539
 DB 479 AQMKAEKWIYAYENWVNDIGLQCGRAQIGKGMWAMPDLMAAAMLEQKIAHPLAGANTA 538
 QY 540 WVPSPATAATLHALVHVQDVAAVOGLAGKERRATIEOLLTIPLAKELAWAPDEIRREEDVN 599
 DB 539 WVPSPATAALHALVHVQDVFAQAELAKRERASVDDILTIPLANTDWTTPSQVQNELDN 598
 QY 600 NCOSILGVVVRWVGCGSKVPDIHDVVALMEDRATLRISQILLANLWLRHGVTSDVRA 659
 DB 599 NAQOILGVVVRWIDQGVGSKVPDINDIGLMDRATLRISQILLANLWLRHGVTSDQVME 658
 QY 660 SLERMAPLVDRQAGDVAYRPMNFDSDIAFLAQELILSGAOPNGYTEPIHLRRRRE 719
 DB 659 SLKRMAPVVDQRQAGDALLYRPLRPSPDFDSNIAFQAAVELVIEGTQKPNGYTEPVLHRRRE 718
 QY 720 FKAR 723
 DB 719 FKAR 722
 RESULT 8
 MASZ RHIL0
 ID MASZ RHIL0 STANDARD; PRT; 721 AA.
 AC Q98DK4;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DT

DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MLR4664
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
CX Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303059;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlsB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC
CC EMBL; AF003004; BAB51267.1; -.
DR HAMAP; MF 00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006283; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn. G; 1.
DR Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 338 338 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 629 629 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 721 AA; 78058 MW; 55376311A1E1BFDF CRC64;

Query Match 63.4%; Score 2416; DB 1; Length 721;
Best Local Similarity 64.4%; Pred. No. 2.8e-150;
Matches 467; Conservative 86; Mismatches 164; Indels 8; Gaps 3;

QY 1 TDRVSVGNLRIRVLYDFVNNELPGTDIDPDSFWAGVGVKVDLTPQCALINARDELQ 60
DB 2 TDRIEAGURIAELHDFVAGEALPGTGAADAFWSGSAIVHDLAPKRNALLKKRDAMQ 61
QY 61 AQIDKWHRRVRIPIIDMDAYRQELTEIGYLLPPDDFTTTSGVDAEITTTAGPOLVVPV 120
DB 62 ERLDGVWRDNG-APDMEVYKSLKEIGLVNPEGPAPFVSTDNVDEIAVAGPQLVVPV 120
QY 121 LNAFALNANANRWGLYDALYGTVDVTPEDGAKGFTYKVRGDKVIAVAKFLDDSV 180
DB 121 MNARYALNANANRWGLYDALYGTVDVTPEDGAKGFTYKVRGDKVIAVAKFLDDSV 180
QY 181 LSSGSPGDATGFTVQDQQLVVALPDKSTGLANPGQFAGYTGAEPSVLLINHLGHIET 240
DB 181 LTSGKAGVNGLSVANGALKLGAGAGTTLADPQFAGVGRGDAANDPADVLLVYNGHLHIEI 240
QY 241 LIDPESQVGTDRAGVKDVILSAITIMDFEDSVAADAAKVLGSRNWLGNKGDZAA 300
DB 241 VDRNNQIGRTDPAGTADVILESAITIQDCEDSVAADAAKVLGSRNWLGNKGDZAA 300
QY 301 AVDKDGTAFILVLRNDRNTAPGGQFTLPGSLMFVNVGHLMNTDAIVDTGSEVPEG 360
DB 301 EISKAGRSFVRKLNADRAYTAPAGGQITVPGSLMLVRNVGHLMNTDAIVDTGSEVPEG 360

QY 361 INDALFTGLIAIHGLKASDV--NGPLNSRTGSIYIVKPKHGHGAFAVFTCELSRVEDV 418
DB 361 INDAAITATIALH-----DVGPEGRANRANRAGSVYVVKPKHGHGAFAVEIEDRVEAL 415
QY 419 LGLPONTMKIGIMDEERTTNLAKACIAAADRVVFINFGDLRTGDEIHTSMAGPMVR 478
DB 416 LGMPENTIKGIMDEERTTNLKEAIPAERVRVFINFGDLRTGDEIHTSMAGPMIR 475
QY 479 KGTMSQPWILAYEDHNVNDAAGLAAGFSGRAQVGKGMWMTMELMADMVETKIAQPRAGAST 538
DB 476 KGDMMQAAWISAYEAWNVDTGLECGLAGHAGIGKGMWAMPDLMAAMLEKQIAHPKAGANT 535
QY 539 AWPSPPTAATHALHYHOVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPDIREVD 598
DB 536 AWPSPPTAATHALHYHVDVHVAQAALKSRPKAKLDDILSVFVAVRPNWTPDIEQRELD 595
QY 599 NNCOSILGVYVRVDQGVGCKVPDIHDVVALMEDRATLRISSQLLANLWLRHGVITSADVR 658
DB 596 NNAQILGVYVRWDQGVGCKVPDINDVGLMEDRATLRISSQHLANLWLRHKVCSEIQVR 655
QY 659 ASLERMAPLVDPRONAGDVAYRPMAPNPDSDTAFLAAQELILSGACQPNGYTEPILHRRRR 718
DB 656 DSLQMAAIVDRQNVGDPFLYRMAPDFDKSIATQAACDLVFKGTSQPNNGYTEPVLHARRL 715
QY 719 EFKAR 723
DB 716 ELKAQ 720

RESULT 9
MASZ AGRTS
ID MASZ AGRTS STANDARD; PRT; 731 AA.
AC Q8U085;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
OS GlsB OR ATU0047 OR AGR_C 78.
OC Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Minks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kuyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.,
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlet K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.,
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58";
RL Science 294:2323-2328(2001).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.

```

CC CC -!- SUBUNIT: Monomer (By similarity).
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008979; AAL41078.1; ALT_INIT.
DR EMBL; AE007947; AAK85871.1; ALT_INIT.
DR PIR; AH2582; AH2582.
DR HAMAP; MF_00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRfam; TIGR01345; malate syn.G. 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT SITE 346 CATALYTIC BASE (BY SIMILARITY).
FT ACT SITE 637 CATALYTIC ACID (BY SIMILARITY).
FT ACT SITE 637 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 731 AA; 79520 MW; 69F304D5D6F8EFB CRC64;
-----
Query Match 62.3%; Score 2372.5; DB 1; Length 731;
Best Local Similarity 63.1%; Pred. No. 4.2e-147;
Matches 461; Conservative 92; Mismatches 165; Indels 7; Gaps 3;
-----
QY 1 TDRVSGNLRIRVLYDFVWNEALPGTDIDPDPSFWAGVKVADLTQNALNARELQ 60
DB 13 TDKP--GLSIDRLIAYFLTVLPGTGLDSETFEGFSAIVHLSPKNRELLAKRDALQ 69
QY 61 AQIDKWHRRVIEPIDMAYQFQLTEIGYLLPEDEPDTTITSGVDALITTAGPOLVVPV 120
DB 70 EKIDGWYENG-APSDPDYAEFLKEIGYLLPEGGPKVETNNVDPEIIVAGPOLVVPV 128
QY 121 LNARALNAANRWGSLYDALYGTVDVTPETDCAKGYPTYNKVRGDKVIAYARFLDSDVP 180
DB 129 MWARALNAANRWGSLYDALYGTDAISDAGAKRGYNPKGDKVIAYARFLDSDAP 188
QY 181 LSSGFGDATGFTVDDGQLVVALPDKSTGLANPQFAGYTCGAASPTSVLLINHLHIEI 240
DB 189 LETGSWSDVTGFTNADGLQLAIGAAATGLKDAVQFKGFGSEAAKPAATILGKNGLHTEI 248
QY 241 LIDPSQVGTTRAGVKDVIIESATITIMDFEDSVAADADKVLGYRNMLGNKGLAA 300
DB 249 VIDPSTGKGRAGISVIESATITIMDCEDSVAADADKVLGYRNMLGNKGLDTE 308
QY 301 AVDKDGTAFRLVLRNDRNTYAPGGQFTLPGRSLMVRNVGHMTNDAIVDTGDSVFEG 360
DB 309 AVSKGNTFTRELNDRYVTAPDGSALTLPGRSLMVRNVGHMTNDAIVDTGDSVFEG 368
QY 361 INDALFTGLIAHGLKASDVNGPLNSPTGSIYTVKPKHGPAEVAFTCELSFVEDVLG 420
DB 369 INDVAVTALIALYDVGPS--GRRNSRAGSMVYVKKPKHGPEEVAFAVELFVENVLG 425
QY 421 LPQNTWKIGIMDEERTTNLAKACIAAADRIVFINTGFLDRTGDEIHTSMEAGPMVRKG 480
DB 426 MAPNTMKGIMDEERTTNLAKESIRAAKDRVVFINTGFLDRTGDEIHTSMEAGPMVRKG 485
QY 481 THKSQFWILAYEDHNVDAGLAAGFSGRAQVKGKMTWTELADNMEVETKIAOPRAGASTAW 540
DB 486 DMKQAAWIAAYENWNVDTGLECLSGSHAQIGKGMWAMPDLMAALMEQKIAHPKAGANTAW 545
QY 541 VPSPTAATHALHYHQVDVAAVQOQGLAGKRATIEQLLTIPLAKELAWAPDEIEEVDNN 600
DB 546 VPSPTAATHATHYHKVDVAAVQOGLKSGRAKLSDIILSVPAFPNWTPEEQLELNN 605
QY 601 CQSILGYVVRVWDQGVGSKVPDIDHVALMEDRATLRISQAHMANWLRHGVVTEAQI 660

```

```

DB 606 AQGILGYVVRVWDQGVGSKVPDINNIGLMEDRATLRISQAHMANWLRHGVVTEAQI 665
QY 661 LERMAPLYDRQAGDVAYRMAPNFDDSIAPLAAQELTSLGAQOPNGVTEPILHRRREF 720
DB 666 MKRMAVVDTONAGDPAYLPMAISDFDGSVAFQAAVELVLKRGQFNGTPEVLRRLLE 725
QY 721 KARAA 725
DB 726 KAKQA 730
-----
RESULT 10
MASZ RHILV STANDARD; PRT; 723 AA.
AC Q937M7;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MASZ.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF39;
RA Garcia de los Santos A., Hynes M.F.;
RT "Malate synthase gene from Rhizobium leguminosarum.";
RL Submitted (OCR-2001) to the EMBL/GenBank/DBSJ databases.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H2O + glyoxylate = S-malate +
CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY059637; AAL17965.1; -.
DR HAMAP; MF_00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRfam; TIGR01345; malate syn.G. 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
FT ACT SITE 338 CATALYTIC BASE (BY SIMILARITY).
FT ACT SITE 629 CATALYTIC ACID (BY SIMILARITY).
FT ACT SITE 629 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 723 AA; 79677 MW; 4E879906CFD6444 CRC64;
-----
Query Match 62.1%; Score 2367.5; DB 1; Length 723;
Best Local Similarity 64.1%; Pred. No. 4.2e-147;
Matches 464; Conservative 82; Mismatches 173; Indels 5; Gaps 3;
-----
QY 3 RVSGNLRIRVLYDFVWNEALPGTDIDPDPSFWAGVKVADLTQNALNARELQAO 62
DB 3 RVDKNGLAETVLDHPLVEEVLPLGLAVDAKDFSAIVHDLAPKNCALLAKRDQLVK 62
QY 63 IDKWHRRVIEPIDMAYQFQLTEIGYLLPEDEPDTTITSGVDALITTAGPOLVVPV 122
DB 63 IDWYRRHG-APADNDEYQSFLEIGYLLPEGSDQVSTQNVDPFIASITAGPOLVVPV 121
QY 123 ARPALNAANRWGSLYDALYGTVDVTPETDCAKGYPTYNKVRGDKVIAYARFLDSDVPLS 182
DB 122 ARVALNAANRWGSLYDALYGTDAIPESDCAEKGSYNPKRGKVIAYARFLDSDAPLQ 181
QY 183 SGSGFGDATGFTVDDGQLVVALPDKSTGLANPQ- FAGYTGAESPTSVLLINHLHIEI 241

```

Db 182 DCRWKDVGSAVKDQALVRSIDGQAWLTDGKHFGYRGDAAPATHILKNNGHIIEIV 241
Qy 242 IDPESQVGTTRAGVKDQVILESAITTIMDFEDSVAADKVKLVGNWGLKNGDLAAA 301
Db 242 IDAATTIGKADSAHISDVWLESAITTIMDCEDSIAAADAEDKVVVRNVLGMLKNGDLOEE 301
Qy 302 VDKDGTAFRLVNRDRNTYAPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDGSVEFEG 361
Db 302 VAKGTSFIRTLNPDLOYAGPDGAFAEVHRRSLMLVRNVGHLMTNPAILLDRDQNEVPEGI 361
Qy 362 MDALFTGLIAHGLKASDVNGPLNSHTGSIYIVKPKMHGPAEVAFTCELSFVSVEDVLGL 421
Db 362 MDAAITGLIAHGLKASDVNGPLNSHTGSIYIVKPKMHGPAEVAFTCELSFVSVEDVLGL 418
Qy 422 PONTMKGIMDEERRITVNLKACIKAAADRWVFINTGFLDRTGDEIHTSMEAGPMVRKGT 481
Db 419 PRNTIKMGIEMDEERRITVNLKACIKAAADRWVFINTGFLDRTGDEIHTSMEAGPMVRKGT 478
Qy 482 MKSQPWTLAYEDHNVDAGLAAGSRAQVKGKGMWMTLMADVETKIAOPRAGASTAW 541
Db 479 MROAWTSAYENWNVNVDIGLECGLAGHAQIGKGMWMPDLMAAMLEQKIAHPKAGANTAW 538
Qy 542 PSPTAATLHALYHVDVAQVQGLAGKRRATIEQLTITPLAKELAWAPDEIRVEVDNCC 601
Db 539 PSPTAATLHALYHVRNVNVAQVQGLKDRAPAKLSDILSVPAVRPNTWTEIEIQRELDNNA 598
Qy 602 QSLGIVYVRVVDQGVGSKVPDIDHVALMEDRATLRISSOLLANWLRHGVITSADVRASL 661
Db 599 QGILGYVVRVVDQGVGSKVPDINNVLGMDRATLRISSACHMANWLRHGVITSADVRASL 658
Qy 662 ERMAPLVDRONAGVAYRPMAPNFDSDIAFLAAGELILSGAOPNGYTPFILLHRRRREF 721
Db 659 RMAAVVDRONAGVAYRPMAPNFDSDIAFLAAGELILSGAOPNGYTPFILLHRRRREF 718
Qy 722 ARAA 725
Db 719 AKQA 722

RESULT 11
MASZ_RHIME
ID MASZ_RHIME STANDARD; PRY 723 AA.
AC Q2TR4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GUCB OR R00062 OR SMC02581.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Fuenler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -/- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -/- PATHWAY: Glyoxylate bypass; second step.
CC -/- SUBUNIT: Monomer (By similarity).
CC -/- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -/- SIMILARITY: Belongs to the malate synthase family. GUCB subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL591782; CAC41449.1; -;
DR HAMAP; MF 00641; -; 1.
DR InterPro; IPR001465; Malate_synthase.
DR InterPro; IPR006253; Malate_synthase.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; Malate_syn_G; 1.
DR Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITES 338 338 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 629 629 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 723 AA; 78853 MW; A0E95B8A5164B58 CRC64;

Query Match 61.7%; Score 2351.5; DB 1; Length 723;
Best Local Similarity 62.9%; Pred. No. 4.6e-146;
Matches 455; Conservative 94; Mismatches 169; Indels 5; Gaps 3;

QY 2 DEVSQGNLRIRARLVYDFVNNALPGTDIDPDSFVAGVDKVVADLTTPQNALINARDELOA 61
Db 2 DRVEKYQIDAGLHRLVFEESAMPFGTGVDAORFFSFDLDVHDLGPKNRALLVXRDELOA 61
QY 62 QIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAAETTTAGPOLVVPVL 121
Db 62 RLDGVYREHG-APVDMEAYEAFLEIGYLLPEGDPFYVSTANVSEIATIAGPOLVVPVM 120
QY 122 NARFALNANARWGLYDALYGTVDVPTDGAEGKPTNKVGRDKVIAYAKFLDSDVPL 181
Db 121 NARYALNANARWGLYDALYGTDAIETDGAERKGYNPKRGAKVIAWAREFLDASAPL 180
QY 182 SSGSGDGTATFTVDGQLVVALPD-KSTGLANPQFAGYTGAAESPTSVLLINLGLHIEI 240
Db 181 AAGRWSDAKSFVSGATLTVTLADTKSAPRNSVQFAGYAGDPAAPSEIVLRRLGLHVI 240
QY 241 LIDPESQVGTTRAGVKDQVILESAITTIMDFEDSVAADKVKLVGNWGLKNGDLAAA 300
Db 241 VLDATTPIGKADAAGISDVWLESAITTIMDCEDSIAAADAEDKVVVRNVLGMLKNGDLEE 300
QY 301 AVDKDGTAFRLVNRDRNTYAPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDGSVEFEG 360
Db 301 EVTKGGAFTRLNPDLOYAGPDGAFAEVHRRSLMLVRNVGHLMTNPAILLDRDQNEVPEG 360
QY 361 IMDALFTGLIAHGLKASDVNGPLNSHTGSIYIVKPKMHGPAEVAFTCELSFVSVEDVLG 420
Db 361 LMDAVVTALIALHDIGR---NGRRANSRSGSVVVKPKMHGPEEVAFACEIFARVEAALG 417
QY 421 LPONTMKGIMDEERRITVNLKACIKAAADRWVFINTGFLDRTGDEIHTSMEAGPMVRK 480
Db 418 LPANAMKMGIMDEERRITVNLKACIKAAADRWVFINTGFLDRTGDEIHTSMEAGPMVRK 477
QY 481 TMKSQPWTLAYEDHNVDAGLAAGSRAQVKGKGMWMTLMADVETKIAOPRAGASTAW 540
Db 478 DMKQAPWTSAYENWNVNVDIGLECGLAGHAQIGKGMWMPDLMAAMLEQKIVHPKAGANTAW 537
QY 541 VPSPTAATLHALYHVDVAQVQGLAGKRRATIEQLTITPLAKELAWAPDEIRVEVDN 600
Db 538 VPSPTAATLHALYHVRNVNVAQVQGLAGKRRATIEQLTITPLAKELAWAPDEIRVEVDN 597
QY 601 CQSLGIVYVRVVDQGVGSKVPDIDHVALMEDRATLRISSOLLANWLRHGVITSADVRAS 660
Db 598 AQGILGYVVRVVDQGVGSKVPDINNVLGMDRATLRISSACHMANWLRHGVITSADVRAS 657
QY 661 LERMAPLVDRONAGVAYRPMAPNFDSDIAFLAAGELILSGAOPNGYTPFILLHRRRREF 720
Db 658 MORMAAVVDGONAGVAYRPMAPNFDSDIAFLAAGELILSGAOPNGYTPFILLHRRRREF 717
QY 721 KAR 723
Db 719 AKQA 722

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE014458; AAN30550.1; --

TIGR; BR1648; --

HAWAP; MF 00641; --

InterPro; IPR001465; Malate synthase.

InterPro; IPR006253; Malate synthase.

Pfam; PF01274; Malate synthase; 1.

TIGRFAMs; TIGR01345; malate syn G; 1.

Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;

Complete proteome.

ACT_SITE 345 CATALYTIC BASE (BY SIMILARITY).

ACT_SITE 636 CATALYTIC ACID (BY SIMILARITY).

SEQUENCE 728 AA; 79966 MW; F95869D002A14EDE CRC64;

Query Match 60.7%; Score 2313; DB 1; Length 728;

Best Local Similarity 61.6%; Pred. No. 1.5e-143;

Matches 443; Conservative 97; Mismatches 175; Indels 4; Gaps 2;

QY 4 VSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVKVADLTQNALNARDELQAOI 63

Db 10 VEIEGLAVAPELVEFLAKEAPGTGVEPEKFKGFAAIRDLAPKRALIAKRDQLQARI 69

QY 64 DKWHRRIEPIIDMDAYRQFLTEIGYLLPEDDFTITTSVDAEIITTAGPQLVPLVNA 123

Db 70 DANYKENRDKGYSCADYQQLKIDIGYLLPEGGAFSVSTINVDPEITHIAGPQLVPLVMA 129

QY 124 RFALNANARWGLYDALYGTVDIPETDGAEGKPTYNKVRGDKVIAIYARKFLDSDVPLSS 183

Db 130 RYALNANARWGLYDALYGTDAISEADGAEGKGNPKRGEKVIAWAKNFLDESAPLST 189

QY 184 GSGDAGTFTVQDQGLVVALPDKS-TGLANPGFAGYTGNAESPTSVLLINHLHIEILL 242

Db 190 GKWADVAGLANVQKLEIRITDGSATLKDDESQFKYNGDAASPTVLLIAKRNHVDIVI 249

QY 243 DPESQVGTTRAGVKDQVILESAITTIMDFSDVAADKVLGYRNWGLNKGDLAAAV 302

Db 250 NADHPKIGTDPADIAVWLESASTIODCEDSAVDAEDKVAVYRNWGLNKGLEDTF 309

QY 303 DKDGTAFRLVNRDRNYTAPGGQFTLPGRSLNFVNRVGHLMNTDAIVTDGSEVPEGIM 362

Db 310 EKNGKQNTRELNGDRITYTAPDGSFTLKGHSLMLVRNVGHLMNTNPAILDAGNEVPEGIM 369

QY 363 DALFTGLIATHGKASDVNGPLNSRTGSIYIVKPKMGHPAEVAFTELFPSVEDVLGP 422

Db 370 DAFTSLIALHDI---GPNGRHNNSREGSVYIVKPKMGHPAEVAFNEFTITEMLGK 426

QY 423 QNTMKIGIMDEERTTNLAKIAAADRVPVINTGFLDRTGDEIHTSMEAGPMVRKGTM 482

Db 427 PNTLKIGIMDEERTTNLKEATRAAKDRVPVINTGFLDRTGDEIHTSMEAGPMVRKGTM 486

QY 483 KSQPTILAYEDHNVDAGLAAGFSGRQVKGGMWMTTELKADNVETKIAOPRAGASTAWVP 542

Db 487 KQANWIGAVEQWVDIGLEGLSGHAQIGKGMWMPDMWMAAMLEQKIAHPRAGASTAWVP 546

QY 543 SPTAATLHLYHQVDVAACQGLAGKRATIEQLLTIPLAKELAWAPDEIEEVDNCCQ 602

Db 547 SPTAATLHLYHQVDVAACQGLAGKRATIEQLLTIPLAKELAWAPDEIEEVDNCCQ 606

QY 603 SILGYVVRWVQGVGSKVPDIHDVALMEDRATLRISSOLLANLWHLGHVITSADVRASLE 662

Db 607 GILGYVVRWVQGVGSKVPDIHDVALMEDRATLRISSOLLANLWHLGHVITSADVRASLE 666

QY 663 RMAPLVDRQVAGVYRPNAPNPDSDIAFLAQELILSGAQQPNGYVTEPILHRRRREFK 721

Db 667 RMAAIVDKQNEGDPLRPMADFDKSIATQACCDLVFKREQPNGYVTEPILHRRRLELK 725

RESULT 14

MASZ_CORGL

ID MASZ_CORGL STANDARD; PRT; 738 AA.

P42450;

01-NOV-1995 (Rel. 32, Created)

01-NOV-1995 (Rel. 32, Last sequence update)

15-SEP-2003 (Rel. 42, Last annotation update)

Malate synthase G (EC 2.3.3.9).

GLCB OR ACEB OR GCL329.

Corynebacterium glutamicum (Brevibacterium flavum).

Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;

Corynebacterineae; Corynebacteriaceae; Corynebacterium.

NCBI TaxID=1718;

OX

[1] SEQUENCE FROM N.A., SEQUENCE OF 1-10. AND CHARACTERIZATION.

RP STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;

RC MEDLINE=95111631; PubMed=7812449;

RX Reinscheid D.J., Eikmanns B.J., Sahm H.;

RA "Malate synthase from Corynebacterium glutamicum: sequence analysis of the gene and biochemical characterization of the enzyme.";

RT Microbiology 140:3099-3108(1994).

RL

[2] SEQUENCE FROM N.A.

RP STRAIN=ATCC 13059 / AS019;

RC Lee H.S., Sinskey A.J.;

RT "Molecular characterization of aceB, a gene encoding malate synthase in Corynebacterium glutamicum.";

RT J. Microbiol. Biotechnol. 4:256-263(1994).

RL

[3] SEQUENCE FROM N.A. / DSM 20300 / NCIB 10025;

RP STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;

RC Nakagawa S.;

RA "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";

RT Submitted (May-2002) to the EMBL/GenBank/DBJ databases.

RL

CC -! CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate + CoA.

CC

CC -! ENZYME REGULATION: Inhibited by oxalate, glycolate and ATP.

CC

CC -! -! PATHWAY: Glyoxylate bypass; second step.

CC

CC -! -! SUBUNIT: Monomer.

CC

CC -! -! SUBCELLULAR LOCATION: Cytoplasmic.

CC

CC -! -! SIMILARITY: Belongs to the malate synthase family. GICB subfamily.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC

CC EMBL; X78491; CAA55243.1; --

DR EMBL; L27123; AAA68074.1; --

DR EMBL; AP005281; BAB99722.1; --

DR PIR; I40715; I40715.

DR HSP; P37330; I08C.

DR HAWAP; MF 00641; --

DR InterPro; IPR001465; Malate synthase.

DR InterPro; IPR006253; Malate synthase.

DR Pfam; PF01274; Malate synthase; 1.

DR TIGRFAMs; TIGR01345; malate syn G; 1.

DR Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;

KW Complete proteome.

FT INIT MET C

FT ACT_SITE 355 CATALYTIC BASE (BY SIMILARITY).

FT ACT_SITE 646 CATALYTIC ACID (BY SIMILARITY).

SQ SEQUENCE 738 AA; 82231 MW; 72AA0663AE7C87E4 CRC64;

Query Match 60.1%; Score 2289; DB 1; Length 738;

Best Local Similarity 61.3%; Pred. No. 5.8e-142;

Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;

QY 1 TDRVSGNLRIRARVLYDFVNNALPGTDIDPDSFWAGVKVADLTQNALNARDELQ 60

Db 19 TERVDAGGMQVAKVLYDFVTEAVLPRVGVDAEKWSGFAALARDLTFRNRELLARRDELQ 78

QY 61 AQIDKWHRRRIEPIIDMDAYRQFLTEIGYLLPEDDFTITTSVDAEIITTAGPQLVPLV 120

```

DB 79 MLIDYHNN-SGIDQEAEDFLKEIGYLVEPEAAEINTQNDVTEISDTAGPOLVPI 137
QY 121 LNARFALNAANRWGSLYDALYGTVDVITPDGAEKGYTNKRGDKVIAVAKPLDSDVP 180
DB 138 LNARFALNAANRWGSLYDALYGTNAIPETDGAEGKEYNPVRGQKVIWEGREFLSDVP 197
QY 181 LSSGSGDATGFTVQDGLVVALPKSTGLANPQAFAGYTGAAESPVSLLINHLHIEI 240
DB 198 LGASHADVEKYNITDGLAAHIGDSVRLKNRESYRGFTGNFLDPAIFLLENGHLIEL 257
QY 241 LIDPESQVGTTRDAGVKDVLIESAITTIMPDFSVAADAAKVLGYRNVLGNKGLDAA 300
DB 258 QIDPWHPIGKADKTGLKDVLESAITTIMPDFSVAADAAEDKTGLYSNNFGLNTGSLKE 317
QY 301 AVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDALIVDTGSEVFE 360
DB 318 EMSKNGRIFTBELAKDRYVIGRNGTELVLHGRSLFLVRNVGHLNQNPSIL-IDGEEIFEG 376
QY 361 IMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMGPAEVAFTCELFGRVEDVLG 420
DB 377 IMDAVLTTCVCAIPGIAPON---KXNSRKGSIYIVKPKMGPAEVAFTCELFGRVEDL 433
QY 421 LPQNTMKIGIMDEERTTNNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKG 480
DB 434 LPRHTLKVGWVDEERTTNNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKA 493
QY 481 TMKSQPIWILAYEDHNVDAAGLAAFGSGRAOVGKGMWMTLMADVETKIAQPRAGASTAW 540
DB 494 DMQTAFWKQAVENNVNDAGIORGLPKAQIGKGMWMTLMADVETKIAQPRAGASTAW 553
QY 541 VPSPTAATLHALHYHQVDVAQVQGL-AGKRATIEQLLTIPLAKELAWAPDEIREVDN 599
DB 554 VPSPTGATLHATHYHLVDVFKVQDELRAAGRRDRLNLTPTAPNTNWSSEKKEMDN 613
QY 600 NCOSILGYVVRWVDOGVGCSKVPDTHDVALMEDRATLRISQLLANLHGHVITSADVRA 659
DB 614 NCOSILGYVVRWVDOGVGCSKVPDTHDVALMEDRATLRISQLLANLHGHVITSADVRA 673
QY 660 SLERMAPLVDRQNAQDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRE 719
DB 674 SLERMAPLVDRQNAQDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRE 733
QY 720 FKAR 723
DB 734 FKAR 737

```

RESULT 15

```

ID MASZ_COREF STANDARD; PRT; 748 AA.
AC Q8FNB3;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MASZ OR CE2231.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -! PATHWAY: Glyoxylate bypass; second step.
CC -! SUBUNIT: Monomer (By similarity).

```

```

CC -! SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -! SIMILARITY: Belongs to the malate synthase family. GICB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP005221; BAC19041.1; -.
DR HAMAP; MF_00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn_G; 1.
KW Transferase; glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
KW ACT_SITE 362 362 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 653 653 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 748 AA; 83491 MW; F9550473EC4E9A09 CRC64;

```

```

Query Match 59.7%; Score 2273; DB 1; Length 748;
Best Local Similarity 61.5%; Pred No. 6.5e-141;
Matches 444; Conservative 90; Mismatches 182; Indels 6; Gaps 4;
QY 1 TDRVSQVGNLRIARVLYDFVNEALPCTDIDPDSFWAGVDKVVADLTPQNALLNARDLQ 60
DB 26 TERVTVGGVQVAKVLRDLFTESVLPKRVGDAERFWNGFGDIVRDMTPRRELLARDELQ 85
QY 61 AQIDKWHRRVTEPDMDAYRQFLTEIGVLLPEPDDFTTTSGVDABITTAGPOLVVPV 120
DB 86 AOLDEYRENPCKP-DPEKYEAFLEIGLVDEPAPAIRQNIQDSEIATTAGPOLVVPV 144
QY 121 LNARFALNAANRWGSLYDALYGTVDVITPDGAEKGYTNKRGDKVIAVAKPLDSDVP 180
DB 145 LNARFALNAANRWGSLYDALYGTNAIPDDEGAERGAEYNPVRGQKVIWEGREFLDVA 204
QY 181 LSSGSGDATGFTVQDGLVVALPKSTGLANPQAFAGYTGAAESPVSLLINHLHIEI 240
DB 205 LDGASHADVEKYNITDGLAAHIGDSVRLKNRESYRGFTGNFLDPAIFLLENGHLIEL 264
QY 241 LIDPESQVGTTRDAGVKDVLIESAITTIMPDFSVAADAAKVLGYRNVLGNKGLDAA 300
DB 265 QIDPWHPIGKADKTGLKDVLESAITTIMPDFSVAADAAEDKTGLYSNNFGLNTGELTE 324
QY 301 AVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDALIVDTGSEVFE 360
DB 325 EVAKGDRTETRLNDRDVRVFGKNGAELTLHGASLLFVRNVGHLMTNPAIL-VDGSEIYEG 383
QY 361 IMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMGPAEVAFTCELFGRVEDVLG 420
DB 384 IMDAITTCVCAIPGIAPONKKK---NSRKGSIYIVKPKMGPAEVAFTCELFARVEDLJD 440
QY 421 LPQNTMKIGIMDEERTTNNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKG 480
DB 441 LPRHTLKVGWVDEERTTNNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKA 500
QY 481 TMKSQPIWILAYEDHNVDAAGLAAFGSGRAOVGKGMWMTLMADVETKIAQPRAGASTAW 540
DB 501 DMQTAFWKQAVENNVNDAGIORGLPKAQIGKGMWMTLMADVETKIAQPRAGASTAW 560
QY 541 VPSPTAATLHALHYHQVDVAQVQGL-AGKRATIEQLLTIPLAKELAWAPDEIREVDN 599
DB 561 VPSPTGATLHATHYHLVDVFKVQDELRAAGRRDRLNLTPTAPNTNWSSEKKEELDN 620
QY 600 NCOSILGYVVRWVDOGVGCSKVPDTHDVALMEDRATLRISQLLANLHGHVITSADVRA 659
DB 621 NCOSILGYVVRWVDOGVGCSKVPDTHDVALMEDRATLRISQLLANLHGHVITSADVRA 680
QY 660 SLERMAPLVDRQNAQDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRE 719

```

Db 681 SLERMAVVVEQVAGDPNYLNKAPNFTESVAFQAAEDLILKGTSPAGYTEPILHARRR 740
QY 720 FK 721
Db 741 FK 742

Search completed: November 21, 2003, 16:04:15
Job time : 11.7256 secs

This Page Blank (uspio)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 36.4127 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-2
Perfect score: 3810
Sequence: 1 TDRVSGNLRIRARLVDFVN.....KARAEKAPSDRAGDDAAR 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 23: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_prodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriaph: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2634	69.1	724	2 Q9AE55	Q9AE55 rhodococcus
2	2569	67.4	725	16 Q91636	Q91636 pseudomonas
3	2416	63.4	721	16 Q98DK4	Q98DK4 rhizobium l
4	2374.5	62.3	744	16 Q8UJ95	Q8UJ95 agrobacteri
5	2367.5	62.1	723	2 Q937W7	Q937W7 rhizobium l
6	2351.5	61.7	723	16 Q92TA4	Q92TA4 rhizobium m
7	2318	60.8	728	16 Q8YIR3	Q8YIR3 brucella me
8	2313	60.7	728	16 Q8FZ50	Q8FZ50 brucella su
9	2273	59.7	748	16 Q8FNB3	Q8FNB3 corynebacte
10	2256.5	59.2	727	16 Q9XB03	Q9XB03 bacillus ha
11	2168.5	56.9	723	16 Q8FDN6	Q8FDN6 escherichia
12	279.5	7.3	538	16 Q8FQR6	Q8FQR6 xanthomonas
13	258	6.8	540	16 Q8PDJ8	Q8PDJ8 xanthomonas
14	206	5.4	907	5 Q9U3Q5	Q9U3Q5 caenorhabdi
15	198.5	5.2	968	5 O17353	O17353 caenorhabdi
16	190.5	5.0	533	16 Q8X609	Q8X609 escherichia

```

17 190 5.0 533 16 Q8Z1W0
18 187 4.9 533 16 Q8ZKL4
19 187 4.9 556 16 Q9KTZ6
20 186.5 4.9 539 16 Q8FB63
21 184 4.8 543 5 Q8MYJ5
22 182.5 4.8 551 3 Q9P8O2
23 181.5 4.8 187 16 Q8B8F8
24 179.5 4.7 530 16 Q93J61
25 179 4.7 545 16 Q8DEX5
26 177 4.6 188 16 Q8D3P1
27 175.5 4.6 525 16 Q9A7F8
28 160 4.2 562 10 Q9LZC3
29 154.5 4.1 202 16 Q9KKZ3
30 154.5 4.1 529 16 Q8XZN2
31 154 4.0 540 16 Q9RKU9
32 149 3.9 532 16 Q8ZAR5
33 149 3.9 543 16 Q8DIU4
34 149 3.9 549 16 Q8EGV8
35 147 3.9 540 2 Q9ALX7
36 143.5 3.8 542 3 Q93857
37 142 3.7 528 2 Q8VM95
38 142 3.7 556 10 Q9AXL7
39 140 3.7 813 17 Q8ZXH2
40 139 3.6 3455 16 Q9P9U6
41 138.5 3.6 531 16 Q8ENS2
42 137.5 3.6 4848 2 Q07944
43 135.5 3.6 521 16 Q9RYN3
44 135.5 3.6 3442 16 Q9PBE8
45 135 3.5 3816 2 Q9KIV3

```

ALIGNMENTS

RESULT 1

```

Q9AE55 PRELIMINARY; PRT; 724 AA.
ID Q9AE55;
AC Q9AE55;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Malate synthase.
GN VICA.
OS Rhodococcus fascians.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1828;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=DI88;
RA Verecke D.M., Cornelis K., Van Montagu M., El Jaziri M., Holsters M.,
RA Goethals K.;
RT "Characterization of a chromosomal locus that affects pathogenicity in
RT Rhodococcus fascians.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301559; CAC35701.1; -.
DR HSSP; P37330; ID8C.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn.G.1.
SQ SEQUENCE 724 AA; 78609 MW; F889F883890995E CRC64;

```

Query Match 69.1%; Score 2634; DB 2; Length 724;
Best Local Similarity 69.0%; Pred. No. 2.2e-162;
Matches 499; Conservative 82; Mismatches 138; Indels 4; Gaps 2;

```

QY 1 TDRVSGNLRIRARLVDFVNNEALPGTDIDPSDWAGVDKVVADLTQONALLNARDLQ 60
   ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 TDRVQAGLQVAKVLDFVEKEALPGTDLSEAFWAGASVIADLAPKALLAVRDIIQ 61
   ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 AQIDKWHRRRVIEPIDMDAYRQFLTEIGYLLPEDDFTITTSVGDAEITTTAGPQLVVPV 120
   :||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Db 62 GKVDHWHGHAEGYDRAAYKAFLEIGYLLDEPADFQIHTSGVDTETITTAGPQLVVPV 121
 QY 121 LNARFALNAANRWGSLYDALYGTDTVPETDGAEGKPTYNKVRGDKVIYARFELDSVP 180
 Db 122 LNARFALNAANRWGSLYDALYGTDAIPETDGAEGKPTYNKVRGDKVIYARFELDSVP 181
 QY 181 LSSGSGFGATGFTVQDGGVVALPKST-GLANPGQFAGYTGAAESPTSVLLINHGHLIE 239
 Db 182 LSSGSHVCTTGVDAAASLTWLDGSIYGLKDSQLLGCQGTDPATAILFVHNGLHFE 241
 QY 240 ILIDPESQVGTDRAGVNDVILESAITIMPFDSVAADKADKVLGYRNWGLKNGDLA 299
 Db 242 IQIDPESPIGKTDGAGVNDVILESAITIMPFDSVAADKADKVLGYRNWGLKNGDLA 301
 QY 300 AAVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSIMFVRNVGHMTNDALVDGSEVFE 359
 Db 302 BEVSKGGKTFTRAMNKRITYTSVDGSELTLLHGRSLLFVRNVGHMTNDALVDGSEVFE 361
 QY 360 GIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMGHGAEPVAFTECELSRVEDVL 419
 Db 362 GILDALFTSLAGLSLTDPNV--LSNSRTGSLYIVKPKMGHGPDEVAFTAELFGRVEQVL 418
 QY 420 GLPQNTWKIGIMDEERTTNLKAICIAAADRVVFINTEGLDRTGDEBIHTSMEAGPVVRK 479
 Db 419 GLPNTLKVGMDEERTTNLKAICIAAADRVVFINTEGLDRTGDEBIHTSMEAGPVVRK 478
 QY 480 GTMKSQPILAYEDHNVNDAAGAFSGRAQVKGKMTMTLMADWVETKIAQPRAGASTA 539
 Db 479 GAMGKXWIAAYEDFNVDTLGAGLQKQKAIKKGWAWPDLMDHLEOKI-GHPKAGANTA 538
 QY 540 WVPSPATATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELANWAPDEIREVDN 599
 Db 539 WVPSPATATLHALHYHQVDVAFQHEIAKAKRATVDEILEPLAPSTDWTDDEEKQNELDN 598
 QY 600 NCQSTLGVVWVWQGVGCKVPDIHDVVALMEDRATLRISQQLANWLRHGVITSADVRA 659
 Db 599 NSQSTLGVVWVWQGVGCKVPDIHDVVALMEDRATLRISQQLANWLRHGVITEQVRE 658
 QY 660 SLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPIHLRRRE 719
 Db 659 SLKRMAPVVDQNASDPTKPLAPDFDNTAFQASDLIFQTSQPNGYTEPIHLRRRE 718
 QY 720 FKA 722
 Db 719 YKA 721

RESULT 2

Q91636 PRELIMINARY; PRT; 725 AA.
 ID Q91636
 AC Q91636
 DT 01-MAR-2001 (TremBLrel. 16, Created)
 DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
 DE Malate synthase G.
 GN GLCB OR PA0482.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAOI;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.B., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goitry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964 (2000).

DR EMBL; AB004485; AAG03871.1; -
 DR HSSP; P37330; LD8C.
 DR InterPro; IPR001465; Malate_synthase.
 DR InterPro; IPR006253; Malate_synthG.
 DR Pfam; PF01274; Malate_synthase; 1.
 DR TIGRFAMs; TIGR01345; malate_synG; 1.
 KW Complete proteome.
 SQ SEQUENCE 725 AA; 78659 MW; 3669670A9E38D391 CRC64;
 Query Match 67.4%; Score 2569; DP 16; Length 725;
 Best Local Similarity 69.1%; Pred. No. 3.7e-158;
 Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;
 QY 1 TDRVSVGNLRIRARLVYDFVNNALPGTDDPDSPWAGVQVADLTTPONQALLNARDLQ 60
 Db 2 TERVQVGGLOVAKVLFDFVNNALPGTGVSAADFTWTGAENVINDLAPKNKALLAKRDELQ 61
 QY 61 AQIDKWHRRRVIEPIMDAYRQFLTEIGYLLPEPDDFTITTSVDAAEITTTAGPQLVVPV 120
 Db 62 AXIDGWHQARQAQAHAVAYRAFLTEIGYLLPEAEDFOAGTQNVDDDEIARMAGPQLVVPV 121
 QY 121 LNARFALNAANRWGSLYDALYGTDTVPETDGAEGKPTYNKVRGDKVIYARFELDSVP 180
 Db 122 MNARFALNAANRWGSLYDALYGTDTVISEEGAGKGYNKVRGDKVIYARFELDSVP 181
 QY 181 LSSGSGFGATGFTVQDGGVVALPKST--TGLANPGQFAGYTGAAESPTSVLLINHGHLIE 239
 Db 182 LSSGSHVDTATSYVNGALVWALXNGSETGLKNAQGLAFQGDAAKPKQAVLLKNGLHFE 241
 QY 240 ILIDPESQVGTDRAGVNDVILESAITIMPFDSVAADKADKVLGYRNWGLKNGDLA 299
 Db 242 IQIDPESPIGKTDGAGVNDVILESAITIMPFDSVAADKADKVLGYRNWGLKNGDLA 301
 QY 300 AAVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSIMFVRNVGHMTNDALVDGSEVFE 359
 Db 302 BEVSKGGKTFTRAMNKRITYTSVDGSELTLLHGRSLLFVRNVGHMTNDALVDGSEVFE 361
 QY 360 GIMDALFTGLIAIHGLKASDVNGPLI--NSRTGSIYIVKPKMGHGAEPVAFTECELSRVED 417
 Db 362 GIDGLFTSLIAIH-----DLNGNTSRKNSRTGSIYIVKPKMGHGAEPVAFTECELSRVED 416
 QY 418 VLGLPQNTWKIGIMDEERTTNLKAICIAAADRVVFINTEGLDRTGDEBIHTSMEAGPV 477
 Db 417 VLGLPNTLKVGMDEERTTNLKAICIAAADRVVFINTEGLDRTGDEBIHTSMEAGPV 476
 QY 478 RKGTMKSQPILAYEDHNVNDAAGAFSGRAQVKGKMTMTLMADWVETKIAQPRAGAS 537
 Db 477 RKGAMKSEKWIAYEDFNVDTLGAGLQKQKAIKKGWAWPDLMDHLEOKIGHPLAGAN 536
 QY 538 TAWPSPATATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELANWAPDEIREV 597
 Db 537 TAWPSPATATLHALHYHQVDVAFQAEAKRTPASVDDIITLPIAPNTNNTAEIKREV 596
 QY 598 DNNQCSILGVVWVWQGVGCKVPDIHDVVALMEDRATLRISQQLANWLRHGVITSADV 657
 Db 597 DNNQCSILGVVWVWQGVGCKVPDIHDVVALMEDRATLRISQQLANWLRHGVITSADV 656
 QY 658 RASLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPIHLRRR 717
 Db 657 VESLXRMAPVVDQNASDPTKPLAPDFDNTAFQASDLIFQTSQPNGYTEPIHLRRR 716
 QY 718 REFKA 723
 Db 717 REFKA 722

RESULT 3

Q98DK4 PRELIMINARY; PRT; 721 AA.
 ID Q98DK4
 AC Q98DK4
 DT 01-OCT-2001 (TremBLrel. 18, Created)
 DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
 DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)

DE Malate synthase G.
GN MR4664.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003004; BAB51267.1; -
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
KW Complete proteome.
SQ SEQUENCE 721 AA; 78058 MW; 65376311A7E1BFDF CRC64;
Query Match 63.4%; Score 2416; DB 16; Length 721;
Best Local Similarity 64.4%; Pred. No. 3.1e-148;
Matches 467; Conservative 86; Mismatches 164; Indels 8; Gaps 3;
QY 1 TDRVSGNLRIRARLVDFYNNALPGTDIDPSPWAGVDKVVADLTPOQALLNARDLQ 60
DB 2 TDRIETAGLRIGELHDFVAGEALPGTGAADAFWSGSAIVHDLAPKRALKKRDAMQ 61
QY 61 AQIDKWHRRVPEIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTTAGPOLVVPV 120
DB 62 ERLDGYRNG-APVDMVEYKSFLEIGYLVPEGPAFVSVDNVPETAWAGPOLVVPV 120
QY 121 LNARFALNANARWGLYDALYGTVDVTPETDGAEGKPTYNKVRGDKVIAYARKFLDDSV 180
DB 121 MARYALNANARWGLYDALYGTDAIPETGGAEGKGNPARGAKVIANAKDFLDQVP 180
QY 181 LSGSGFGDGTFTVQDQGLVWALPDKSTGLANPGPAGVTGAESPTSVLLNHGLHIEI 240
DB 181 LTSKGWAGNGLSVAHGALKLGAAGGTTIADPRPAGYRGDAANPDVALLVKGHLHIEI 240
QY 241 LIDPESQVGTDRAGVKVILSAITTIIDFDSVAADAAKVLGYRWLGLNKGDIAA 300
DB 241 VDRNNQIGRTDAGIADVILSALITIQCDSDVAANDAQKVYVYRWLGLMKGDIAE 300
QY 301 AVDKGTAFRLVLRNDRNTAPGGQFTLPGRSLMFVRNVGHLMNDIAIVDTGSEVFEG 360
DB 301 EISKAGRSFVRKLNADRAVYAPAGGQITVPGRSMLVRNVGHLMNTNPAILDRDGNVEPEG 360
QY 361 IMDALFTGLIAHGLKASDV--NGPLNSRTGSIYVKKPMGPAEVAFTCLFSRVEDV 418
DB 361 IMDAALTALIALH-----DVGPEGRNAGRAGSMYVVKPMGPEEVAFAVIFRVEAL 415
QY 419 LGLPONTMKGIMDEERRITVNLKACIKAAARVVFINTGFLDRDTEHTSMEAGPMVR 478
DB 416 LGMPNTIKGIMDEERRITVNLKEAIRAREVVFINTGFLDRDTEHTSMEAGPMIR 475
QY 479 KGTMKSQPILAYEDHNVDAGLAAGSGRAQVKGKMTWTELMADNVETKIQAPRAGAST 538
DB 476 KGMKQAAMTISAYEAWNVDTGLECGLAGHAQIKGNWAMPDLMAAMLEQKIAPRAGANT 535
QY 539 AWVPSPTAATHALHYHVDVAVQOGLAGKRRATIEQLITPLAKELAWADEIREEVD 598
DB 536 AWVPSPTAATHALHYHVDVAVQOGLAGKRRATIEQLITPLAKELAWADEIREEVD 595
QY 599 NNCQSIGLYVVRWVDQGVCSKVPDIHDVALMEDRATLRISQALLANWLHGHVITSADVR 658
DB 596 NNAQIGILGYVVRWIDQGVCSKVPDINDVGLMEDRATLRISQALLANWLHGHVITSADVR 655

QY 659 ASLERMAPLVDRONAGDVAYRVPAPNPDDSIATIAQELILSGAQPNQNGTETPILHRRR 718
DB 656 DSLQMAAIVDRQNVGLPVRPAPDPKSIATFQAACDLVFKGTQPNQNGTETPVLHARRL 715
QY 719 EFKAR 723
DB 716 ELKQA 720
RESULT 4
Q8UJ85 PRELIMINARY; PRT; 744 AA.
AC Q8UJ85
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Malate synthase G.
GN GLCB OR ATU0047 OR AGR_C 78.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F.Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D.Sr., Grant C.,
RA Chapman P., Dechenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liac L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tinsley S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Houtrel B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlik K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RA "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE008979; AAL41078.1; -
DR EMBL; AE007947; AAK85871.1; -
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
KW Complete proteome.
SQ SEQUENCE 744 AA; 80991 MW; FB74ED3E8D650319 CRC64;
Query Match 62.3%; Score 2374.5; DB 16; Length 744;
Best Local Similarity 62.3%; Pred. No. 1.6e-145;
Matches 462; Conservative 97; Mismatches 162; Indels 21; Gaps 4;
QY 1 TDRVSGNGL-----RIARV-----LYDFVNEALPGTDIDPDSFAGVDKVA 43
DB 6 SDRLSVQSPMEYKEAHVSRTRDKFGLSDRLYAFLTDEVLPGTGLDSETFEGFSAIVH 65
QY 44 DLTQNALNARDLQAQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTS 103
DB 66 ELSPKRELLAKRALQEKIDGWYRENG-APSDFAVEAFLEIGYLLPELPGFKVETNN 124
QY 104 VDAETTTTAGPOLVVPVNLNARFALNANARWGLYDALYGTVDVPEIDGAEKPTYNKVR 163

```

125 VDPBIYAVAGPQLVVPVNNARYALNAANARWGSYDALYGTDAISDADGAEGKRGYNPKR 184
164 GDKVIYARKELDSVPLSSGSGFDATGFTVDDQQLVVALPDKSTGLANPQAGYTGAA 223
185 GDKVIAMARNELDESAPLETSGSDVTGFIADGLQLAIGAAITGLKDAVQKFGSGEA 244
224 ESPTSVLLINHLGLHIEILIDPESQVGTTRAGVADVILESAITTIMDFEDSVAADAAK 283
245 AKPATILGKNGLHTEIVIDPSTIGKSDRAGISDVILESAITTIMDCEDSVAADAEK 304
284 VLGRNWLKNGGLDAAAVDKGTAFILNRNDRNYTAPGGQFTPLPGRSLMFRVNVGHL 343
305 VLVTGNLGLMRGDLTEAVSGKNTFTRLNPDYRYTAPDGSALTPLPGRSLMFRVNVGHL 364
344 MTNDALVDTDCSEVFEGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMHGPA 403
365 MTNPAILDRDQDVEGIMDAVVTALIALYDVGPS---GRQNSRAGSMVIVKPKMHGPE 421
404 EVAFTCELSFVEDVLGPQNTKIGIMDEERTTVNLKACIKAAADRVVFINTEGLDRT 463
422 EVAFAEIFAENVLGVAPNTKMGIMDEERTTVNLKESIRAAKDRVVFINTEGLDRT 481
464 GDEIHTSNEAGPMVKCTKMSQPMILAYEDHNVDAGLAAGFSRAQVKGKMTMTTELMA 523
482 GDEIHTSNEAGPMIRKGMKQAAWIAAYENNVNDIGLECGLSGHAQIGKGMWAMPDLMAA 541
524 NVETKIAOPRAGASTAWPSPATAATLHALHYHOVDVAQVQGLAGKRATIEQLTIPLA 583
542 MLEQKIAHPKAGANTAWPSPATAATLHATHYKVDVAQVQGLKSRGAKLSILSVPA 601
584 KELAWAPDEIREVDNQCOSILGVVVRWDQVGCCKVPDIHVALMEDRATIRISSQL 643
602 PRPNTPEEIQRELDNNAQGLGVVVRWDQVGCCKVPDINNIGLMEDRATIRISAOHM 661
644 ANWLHRGVITSADVRASLERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAQELILSGAQ 703
662 ANWLHRGVVTEAQIIKTMKMNAVVDTQAGDPAYLPWASDFDGSVAFQAAVELVLKRE 721
704 QPNGYTEPILHRRRRREFKAAA 725
722 QPNGYTEPVLHRRRLELKAQA 743

RESULT 5
Q937W7 PRELIMINARY; PRT; 723 AA.
AC Q937W7;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Malate synthase G.
GN MSG.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF39;
RA Garcia de los Santos A., Hynes M.F.;
RL "Malate synthase gene from Rhizobium leguminosarum.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY059637; AAL17965.1;
DR InterPro; IPR001465; Malate_synthase.
DR InterPro; IPR006253; Malate_synthG.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
SQ SEQUENCE 723 AA; 79677 MW; 4E879906CFD64444 CRC64;

Query Match 62.1%; Score 2367.5; DB 2; Length 723;
Best Local Similarity 64.1%; Pred. No. 4.4e-145;
Matches 464; Conservative 82; Mismatches 173; Indels 5; Gaps 3;

```

```

QY 3 RYSGNLRIRVLYDFVNEALPGTDIDPSFAGVQKVVADLTPONALLNARDELQAO 62
DB 3 RYDKNGLAETVLHDFLVEEVLPGLAVDADRFADFAISVHDLAPKNCALLAKDELQOVK 62
QY 63 IDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPVLN 122
DB 63 IDDWYRRHG-APADMDEYQSFLREIGYLLPEGSDQVSTQNVDPETIASIAGPQLVVPVN 121
QY 123 ARFALNANARWGSYDALYGTDPVETDGAEGKPTYNKVGDKVIAVARKELDDSVPLS 182
DB 122 ARYALNANARWGSYDALYGTDAIPESDGAEGKSNPKRGEKVIAWRVDFLDTAPLQ 181
QY 183 SSGSFGDATGFTVDDQQLVVALPDKSTGLANPQ- FAGYTGAESPTS VLLINHLGLHIEIL 241
DB 182 DCRWKDVGSFAVKDGCALVRSIDGQAMLTGKGFAGYRGDAAPTHLLKNGHIEIV 241
QY 242 IDPESQVGTTRAGVQKDVILESAITTIMDFEDSVAADAAKVLGYRNWGLNKGDLAAA 301
DB 242 IDAATTIGKASAHISDVMLESAITTIMDCEDSAAVDAEDKVVYRNWGLMKGDLOBE 301
QY 302 VDKDGTAFILNRNDRNYTAPGGQFTPLPGRSLMFRVNVGHLMTNDALVDTDCSEVFEG 361
DB 302 VAKGTSIRITLNPDLQYAGPDGAFFVHRRSLMFRVNVGHLMTNPAILDRDGNVEPGE 361
QY 362 MDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMHGPAEVAFTCELSFVEDVLGL 421
DB 362 MDAAITGLIALYDIPGS---GRRKNSRTGSMVIVKPKMHGPEEVAFAVEIFSRVEDALGL 418
QY 422 PONTWKIGIMDEERTTVNLKACIKAAADRVVFINTEGLDRTGDEIHTSMEAGPMVRKGT 481
DB 419 PRNTKMGIMDEERTTVNLKCEIRARERVVFINTEGLDRTGDEIHTSMEAGPMVRKGD 478
QY 482 MKSQPMILAYEDHNVDAGLAAGFSRAQVKGKMTMTTELMAVETKTAQPRAGASTAW 541
DB 479 MEQAAMISAYENNVNDIGLECGLSGHAQIGKGMWAMPDLMAALEQKIAHPKAGANTAW 538
QY 542 SPPTAATLHALHYHOVDVAQVQGLAGKRATIEQLTIPLAKELAWAPDEIREVDNQC 601
DB 539 SPPTAATLHATHYKVDVAQVQGLKDRARAKLSILSVVPVAVRPNWTPEEIQRELDNNA 598
QY 602 QSILGVVVRWDQVGCCKVPDIHVALMEDRATIRISSQLANLHRGVITSADVRASL 661
DB 599 QGILGVVVRWDQVGCCKVPDINNIGLMEDRATIRISAOHMANKLHKKVTEQIIETM 658
QY 662 ERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAQELILSGAQCPNGYTEPILHRRRRREFK 721
DB 659 RMAAVVDRQNASDPAYRPMAGNFDDSIAPQALDLVLKRGEPNGYTEPVLHRRRLELK 718
QY 722 ARAA 725
DB 719 AKQA 722

```

RESULT 6

Q92TA4

ID Q92TA4 PRELIMINARY; PRT; 723 AA.

AC Q92TA4;

DT 01-DEC-2001 (TremBLrel. 19, Created)

DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)

DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)

DE Probable malate synthase G protein (BC 4.1.3.2).

GN GLCB OR R00062 OR SMC02581.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Sinorhizobium.

OX NCBI_TaxID=382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,

RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,

[illegible]

ID	Q8YIR3	PRELIMINARY;	PRT;	728 AA.
AC	Q8YIR3			
DT	01-MAR-2002	(T-EMBLrel. 20, Created)		
DT	01-MAR-2002	(T-EMBLrel. 20, Last sequence update)		
DT	01-MAR-2003	(T-EMBLrel. 23, Last annotation update)		
DE	Malate synthase G (EC 4.1.3.2).			
GN	MEI0380.			
OS	<i>Brucella melitensis</i> .			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Brucellaceae; Brucella.			
OX	NCBI_TaxID=29459;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=46X / ATCC 23456 / Biotype 1;			
RX	MEDLINE=20020109; PubMed=11756688;			
RA	DelVecchio V.G., Kapural V., Recker R.J., Patra G., Majer C., Los T.,			
RA	Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,			
RA	Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,			
RA	Ra Selkov E., Blzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,			
RA	Haselkorn R., Kyrpides N. Overbeek R.;			
RT	"the genome sequence of the facultative intracellular pathogen			
RT	<i>Brucella melitensis</i> ."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).			
RL	EMBL; AEO09480; AAL1561.1; -			
DR	InterPro; IPR001465; Malate synthase.			
DR	InterPro; IPR006293; Malate synthase.			
DR	Pfam; PF01274; Malate synthase; 1.			
DR	TIGRfams; TIGR01345; malate_syn_G; 1.			
KW	Lyase; Complete proteome.			
SQ	SEQUENCE 728 AA; 77985 MW; E1313A61797270D CRC64;			
	Query Match	60.8%;	Score 2318;	DB 16; Length 728;
	Best Local Similarity	61.8%;	Prod. No. 7.2e-142;	
	Matches	444;	Conservative	97; Mismatches 174; Indels 4; Gaps 2;
QY	4	VSVGNRIARVLVDFVNNNEALPGTDIDPDSFWGVDKVKVADLTTPQNALNARDELQAQI	63	
DB	10	VEIEGLVAPELVETFAKEAPGTGVEPEKFWKGFAAIIRDLPKPKRALLAKRDELQARI	69	
QY	64	DKHRRRVIEPIDMDAVRQLTETIGYLLPEPDDFTITTSVDAEITTTAGPOLVVPVNA	123	
DB	70	DAMKENRDKGYSQADYQQQLKDIHLLPGGGAFSVSTTVNDEPETHIAGPOLVVPVNA	129	
QY	124	RFALNAANARWGLYDALYGTDIVPETDGAEGKFTYNNKVRGDKVIAYARKFLDDSVPLSS	183	
DB	130	RYALNAANARWGLYDALYGTDAISEADGAEGKGYNPKRGEKVIWAKNFLDSEAPLST	189	
QY	184	GSFGDATGPTVQGOQLVAALPDKS-TGLANPGOFAGYTGAAESPTSVLLNHGLHISILI	242	
DB	190	GKWADVAGLANWDGKLEIRUTDSATLTLDKDESFKGYNGDAASPTNVLLAKNNHVDIVI	249	
QY	243	DPSSQVGTTRAGVKDVILESAITTTIMDFEDSVAADDAADKVLGYRNWLGINKGDLAAY	302	
DB	250	NADHPIGKTDPAHIADVLESALSTIQDCEDSIAAVDAEDKVAYRNWLGANGKLEDTF	309	
QY	303	DKGDTFLVNRDRNTAPGGQFTLPGRSLMFVRNVHGLMTNDAIIVTDGSEVFGIM	362	
DB	310	EKGQKQTRFLNGDRYTPADGSLTLKGRSLMVRNVHGLMTNPAILDRAEGNEVPEGIM	369	
QY	363	DALFTGLIATHGLKASDVNGPLNSRTGSYIVKPKHGPAAEVAFTCELSRVEDVLGLP	422	
DB	370	DAFTSLIALHDI---GPNGRHNSREGSYIVKPKHGPAAEVAFAFANEIITRIEMLGMK	426	
QY	423	QNTWKIGIMDEERTTVNLKACIKAAADRVPVFTNGFLDRTGTDBEHTSMZAGPMVRKGTM	482	
DB	427	PNLIKIGIMDEERTTVNLKEAIRAAKDRVPVFTNGFLDRTGTDBEHTSMZAGPMVRKGTM	486	
QY	483	KSQPWILLAEHDNVNDAAGLAAGFSGRAGVQVGKGMWTMELMADWVETKIAQPRAGASTAWP	542	
DB	487	KQAAMIGAYEQWVNDIGLSCGLSGHAGIQGKGMWMPDMMAANELOKKAHPKAGANTAWP	546	
QY	543	SPTAATLHALHYHQVDVAAYVQGGIAGKRRATIEQLLTIPLAKELAWAPDIREEVNNCQ	602	

Db 547 SPTAATLHATHYHKIDVAQVKLSRPRAKLDDILSPVAVRPNNWTPDDIQHEIDNNAQ 606
 QY 603 SILGVVVRWDQGVGCSKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRASLE 662
 Db 607 GILGVVVRWDQGVGCSKVPDINNVLGMDRATLRISAQHLANWLVHGVVSEAQVMTMK 666
 QY 663 RVAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQBELILSGAQPNQGYTEPIHLRRREFK 721
 Db 667 RVAAIVDKQNEGDPDYRPMADPKSIAFOAACDLVFKGREQPNQGYTEPVLHRRLELK 725

RESULT 8
 Q8FZ50
 ID Q8FZ50 PRELIMINARY; FRT; 728 AA.
 AC Q8FZ50;
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Malate synthase G.
 GN GLCB OR BR1648.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 CX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Unayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraut R., Shetty J., Malek J., van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.B., Lindler L.E., Helling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
 DR EMBL; AE014458; AAN30550.1; --
 DR TIGR; BR1648; --
 KW Complete proteome.
 SQ SEQUENCE 728 AA; 79966 MW; F35669D002A14EDE CRC64;

Query Match 60.7%; Score 2313; DB 16; Length 728;
 Best Local Similarity 61.6%; Pred. No. 1, 5e-141;
 Matches 443; Conservative 97; Mismatches 175; Indels 4; Gaps 2;

QY 4 VSGNLRIRARVLYDFVNEALPGTIDPDSFWAGVDKVVADLTQONQALLNARDELQAI 63
 Db 10 VEIEGLAVAPLVEFLAKAEPGVPEXFWKGFPAIIRDLAPKRRALLKDELQARI 69
 QY 64 DKWRRRRIEPIIDMDAYRQFLTEIGYLLPSPDDPTITTSQVDAEITTTAGPQLVVPVNA 123
 Db 70 DAWYKRDKGYSDAQYQFLKIDIGYLLPEGGAFSVSTINVDPEITHIAGPQLVVPVNA 129
 QY 124 RFALNAANARWGLSYDALYGTVDVPIPETDGAEGKQPTNKVKRGDKVIAYARKFLDSDVPLSS 183
 Db 130 RYALNAANARWGLSYDALYGTDAISEADGAEGKGYNPKGEKVIYAWAKNFLDESAPLST 189
 QY 184 GSPGDATGFTVQDQGLVVALPKDS-TGLANPGQAGYTGAESEPTSVLLINHLHIEILI 242
 Db 190 GKWADVAGLVNDGKLEIRITDGSATTLKDSQFNGDAASPTNVLAKHNMHVDIVI 249
 QY 243 DPESQVETDRAGVKDVLIESAITTIMDFDSVAADVADKVLGYRNWLGNGDLAAAV 302
 Db 250 NADHPIGKTPAHADVVLIESAISTIODCEDSIAADVAEDKVAVYRNWLGNGKLEDITF 309
 QY 303 DKDGTAFELVRNDRNTAPGGGFTLPGRSLFVRNVGHLMNDALVDTDGSSEVEGIM 362
 Db 310 EKNQKQMTRRNGRRTYAPDGSFTLTKHSLMLVRNVGHLMNPAILLDAEGNEVEGIM 369
 QY 363 DALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVLGLP 422
 Db 370 DAFTSLIALHDI---GPNGRHNSRSGSIYIVKPKMHGPEEVAFANEIFTRTEELGMK 426

QY 423 QNTMKIGIMDEBERRTTNVLKACIKAAADRVVFINTGFLDRTGDEIHTSMEAGPMVRKGTM 482
 Db 427 PNTLKIGIMDEBERRTTNVLKAIIRAKDRVVFINTGFLDRTGDEIHTSMEAGPMVRKGTM 486
 QY 483 KSQPMILAYEDHNVDAGLAAGFSQRAQVGKGNWMTMELMADWETKIAQPPAGASTAWVP 542
 Db 487 KQAAAMIGAVEQNWVDIGLECGLSGHAQICGKNWAMPDMMAAAMLEQKIAHPKAGANTAWVP 546
 QY 543 SPTAATLHATHYHVDVAQVQGLAGKRATIEQLITPLAKELAMAPDEIREVDNQCQ 602
 Db 547 SPTAATLHATHYHKIDVAQVKLSRPRAKLDDILSPVAVRPNNWTPDDIQHEIDNNAQ 606
 QY 603 SILGVVVRWDQGVGCSKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRASLE 662
 Db 607 GILGVVVRWDQGVGCSKVPDINNVLGMDRATLRISAQHLANWLVHGVVSEAQVMTMK 666
 QY 663 RVAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQBELILSGAQPNQGYTEPIHLRRREFK 721
 Db 667 RVAAIVDKQNEGDPDYRPMADPKSIAFOAACDLVFKGREQPNQGYTEPVLHRRLELK 725

RESULT 9
 Q8FNB3
 ID Q8FNB3 PRELIMINARY; PRT; 748 AA.
 AC Q8FNB3;
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Malate synthase (EC 4.1.3.2).
 GN MASZ OR CE2231.
 OS Corynebacterium efficiens.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 CX NCBI_TaxID=152794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YS-314 / AT 12310 / DSM 44549 / JCM 11189;
 RA Kawabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
 RA Usuda Y., Sugimoto S.;
 RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005221; BAC19041.1; --
 KW Lyase; Complete proteome.
 SQ SEQUENCE 748 AA; 83491 MW; F9550473EC4E9A09 CRC64;

Query Match 59.7%; Score 2273; DB 16; Length 748;
 Best Local Similarity 61.5%; Pred. No. 6, 3e-139;
 Matches 444; Conservative 90; Mismatches 182; Indels 6; Gaps 4;

QY 1 TDRVSGNLRIRARVLYDFVNEALPGTIDPDSFWAGVDKVVADLTQONQALLNARDELQ 60
 Db 26 TERVTVVGQVAKVLRDLFTESVLPVGVDAERFWNGFGDIVRDMTPNNRELLARRDELQ 85
 QY 61 AQIDKWHRRRIEPIIDMDAYRQFLTEIGYLLPSPDDPTITTSQVDAEITTTAGPQLVVPV 120
 Db 86 AQLDYYRENPCKP-DPEKYAEFLREIGYLVDEPAPAEIRTONIDSEIATTAGPQLVVPV 144
 QY 121 LNARPALNAANARWGLSYDALYGTVDVPIPETDGAEGKQPTNKVKRGDKVIAYARKFLDSDV 180
 Db 145 LNARPALNAANARWGLSYDALYGTNAIPDEDEGAEGAEYNPVRGQVIQWRDFLDVL 204
 QY 181 LSSGSFGDATGFTVQDQGLVVALPKDSTGLANPGQAGYTGAESEPTSVLLINHLHIEI 240
 Db 205 LDGASHADVEXNITDGLAAHVNDGIYLRKDRAYLGFTGYFDPFTSILLQNGHLIEL 264
 QY 241 LIDPESQVGTTRAGVKDVLIESAITTIMDFDSVAADVADKVLGYRNWLGNGDLAA 300
 Db 265 QIDPHTPIGKEDKTLKDIILIESAITTIMDFDSVAADVADKVLGYRNWLGNGDLAA 324
 QY 301 AVDKDGTAFELVRNDRNTAPGGGFTLPGRSLFVRNVGHLMNDALVDTDGSSEVEF 360

Db 325 EVAKGDRTRFKLNDRVFIGKNGAELTLHGRSLLFVRNVGHLMTNPAIL-VDGEEIYEG 383

QY 361 IMDALFTGLTAIHLKASDVNGPLINRTGSIYIVKPKHGPAAVFTCELSRVEDVLG 420

Db 384 IMDAITTVCAIPGAPQKK--NSRKGSIYIVKPKHGPAAVFTCELSRVEDVLG 440

QY 421 LPQNTMKIGIMDEERTTVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMAGPMVRKG 480

Db 441 LPRHTLVGVWDEERTSVNLDACIMEVADLAFINTGFLDRTGDEIHTSMAGPMVRKA 500

QY 481 TMSQPHILAYEDHNVDAAGSFGRAQVKGWMTMELMADVETKIAQPRAGASTAW 540

Db 501 DMQAPKQAYEDHNVDAAGSFGRAQVKGWMTMELMADVETKIAQPRAGASTAW 560

QY 541 VPSPTAATLHALHYHVDVAAVOOGL-AGKRRATIEQLITIPAKELAWAPDEIREVDN 599

Db 561 VPSPTGATLHATHYHVRVDVFKVQDELRAAGRRDSLGKILDVPVAPNTNNTDAEKRELDN 620

QY 600 NCOSILGYVVRVVDQVGCQSKVPDIHDVALMEDRATLRSSQILANLWLRHGVITTSADVRA 659

Db 621 NCOSILGYVVRVVDQVGCQSKVPDIHDVALMEDRATLRSSQILANLWLRHGVITTEQVIE 680

QY 660 SLERMAPLVDRQAGDVAVRPMAPNFDSDTAFLAAQELILSGAQCPNGYTERPILHRRRE 719

Db 681 SLERMAVVDEQAGDPVNLNAPNTESVAFQARDLILKGTSFAGYTEPILHARRRE 740

QY 720 FK 721

Db 741 FK 742

RESULT 10

Q9KB03 PRELIMINARY; PRT; 727 AA.

ID Q9KB03

AC Q9KB03

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Malate synthase.

GN BH2133.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RX MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis."

RL Nucleic Acids Res. 28:4317-4331 (2000).

DR EMBL; AP004514; BAB05852.1; -

DR HSSP; P37330; 1D8C.

DR InterPro; IPR001465; Malate synthase.

DR InterPro; IPR006253; Malate synthase.

DR Pfam; PF01274; Malate synthase; 1.

DR TIGRFAMs; TIGR01345; malate_syn; 1.

KW Complete proteome.

SQ SEQUENCE 727 AA; 81208 MW; CE18B45930B5EAE5 CRC64;

Query Match 59.2%; Score 2256.5; DB 16; Length 727;

Best Local Similarity 60.7%; Pred.No. 7.1e-138;

Matches 439; Conservative 95; Mismatches 184; Indels 5; Gaps 3;

QY 6 VGNLRIARLVDFVNEALPGTDIDPDSFWAGYKVDVADLTQNALNARDELQAQIDK 65

Db 7 VGNLQVATKLITIEQALPGTGKIDFWSGVQVQITELMPENKMLAKREEIQATIDA 66

QY 66 WHERRVIEPDMDAYQELTEIGYLLPEPDDFTITTSYGDATITTAGPOLVVPVNLARF 125

Db 67 WHORNK-GPIDFSAYHSFLEEIGYLEPIPEHVITTTTENVDDIEAQAQPOLVVPVNNARY 125

QY 126 ALNANAENWGSILYDALYCTDVIPIETDGAEGKPTYNKVRGDKVIAYARFELDDSVPLSSGS 185

Db 126 AINAANAENWGSILYDALYGSNVSIEDGCEKGTGTYNPKRGTKVIFPAKDFLDTHTFPLTSGS 185

QY 186 FQDAFTGFTVQOQLVALPD-KSTGLANPGQFAGVTGAESPTSVLLINHLGHLEILDP 244

Db 186 HHEALNAYIMDKQLVWVLESKQWTLKDETOFVGQSQGDPSSVILLHLHGLHVEIQIDA 245

QY 245 ESQVOTTRAGVKOVILBSAITTMDPDSVAADADKVLGYRNWGLNGLKGDAAAVDK 304

Db 246 RHPIGKSDRAKQIVLESALTITMDCDSVAADAEKVAVYQNWGLMGKGTLEATFTK 305

QY 305 DQTAFLRYLNRDNYTAPGGQFTLPGRSLMFVRNVGHLMNTDAIVTDGSEVEPEGINDA 364

Db 306 EGKTKRKLINEDRSTAPNGETFSLPGRSLMFVRNVGHLMNTPIRTOSGEVEPEGIDG 365

QY 365 LFTGLIAIHGLKASDVNGPLINRTGSIYIVKPKHGPAAVFTCELSRVEDVLGPN 424

Db 366 IVTSLIAKHDLQ--NGTFRNSKKGSYIVKPKHGSSEVAFANRLFNRIEDILGERH 422

QY 425 TMKIGIMDEERTTVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMAGPMVRKGTNS 484

Db 423 TLKIGWDEERTSVNLKACIEKVKRVVFIINTGFLDRTGDEIHTSMAGPMVRKGTNS 482

QY 485 QFWILAYEDHNVDAAGSFGRAQVKGWMTMELMADVETKIAQPRAGASTAWVPS 544

Db 483 SSMLSAYERSNVAAGLTCGFGQRAQVKGWMTMELMADVETKIAQPRAGASTAWVPS 542

QY 545 TAATLHALHYHVDVAAVOOGLAGKRRATIEQLITIPAKELAWAPDEIREVDNCOGI 604

Db 543 TAATLHALHYHVRVPAIKTLDQTCYRKEIIEIPLAGSTDWNTNEDIQALNNNAQGI 602

QY 605 LGYVVRVVDQVGCQSKVPDIHDVALMEDRATLRSSQILANLWLRHGVITTSADVRLERM 664

Db 603 LGYVVRVVDQVGCQSKVPDIHDVALMEDRATLRSSQILANLWLRHGVITTSADVRLERM 662

QY 665 APLVDRQAGDVAVRPMAPNFDSDTAFLAAQELILSGAQCPNGYTERPILHRRREFKARA 724

Db 663 AKWDEQAGDPVAVRPMADNLEQSVAFQALVELVKGTEQPSGYTERPILHRRREFKOKI 722

QY 725 AEK 727

Db 723 AKE 725

RESULT 11

Q9FDN6 PRELIMINARY; PRT; 723 AA.

ID Q9FDN6

AC Q9FDN6

DT 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Malate synthase G (EC 4.1.3.2).

GN GLCB OR C3705.

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O6.H1 / CFT073 / ATCC 700928;

RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Rosesch P.,

RA Raeko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Ferna N.T.,

RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;

RT "Extensive mosaic structure revealed by the complete genome sequence

RT of uropathogenic Escherichia coli."

RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

DR EMBL; AE016786; AAN82151.1; -

KW Lyase; Complete proteome.

SQ SEQUENCE 723 AA; 80440 MW; AAF740E5FE038F6F CRC64;

Query Match 56.9%; Score 2168.5; DB 16; Length 723;
Best Local Similarity 58.4%; Pred. No. 3.6e-132;
Matches 422; Conservative 113; Mismatches 173; Indels 15; Gaps 5;
QY 4 VSVGNLRIARLVYDFVNEALPGDIDPDSFWAGVGVVADLPQNALNARDELQAOI 63
DB 5 ITQGLRIDANFKRFVDEEVLPGVELDAAFWNNVDEIVHDLAPENRQLLAERDRQAL 64
QY 64 DKWRRRVEIDMAYRQELTEGYLLPEPDDFTTSGVDAEITTTAGPQVWPVNA 123
DB 65 DEWHRSPGPKVAKYKFLREUGYLVPOPHVTVTTGIDSEITTSQAGPQVWPANNA 124
QY 124 RFALNAANARWGSYDALYGTDTVTPETDGAEGKPTNKVGVGDKVIAYARKFLDSDVPLSS 183
DB 125 RYALNAANARWGSYDALYGSIIIPQEGAMVSG--YDQGEQVIAVRFLDESIPLEN 182
QY 184 GSFGDGTFTVQDQVVALPD-KSTGLANPGFAGYTGAESPTSVLLINHGHLIEILI 242
DB 183 GSYQDVAFKVDKQLRIQLKNGKETTILRTPAQVGVGDTAAPTCILLKXNGHLIELQI 242
QY 243 DPESQVGTDRAGVKDVILSAITTIMDFEDSVAAVDAADKVLGYRNWLGKGLAAV 302
DB 243 DANGRIGKDDSAHINDIVEAAISTILDCEDSVAAVDAEDKILLYRNLLGLMGQTLQERM 302
QY 303 DKDGTAFRLVLRNDRNTAPGGGFTLPGRLMFRVNVGHMTNDALVDTDGSEVEGIM 362
DB 303 EKNGRQIVKLRNDROVTAADGSEISLHGRSLLEFIRNVGHMTIPVINDSEGNEIPEGIL 362
QY 363 DALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLGLP 422
DB 363 DGVMTGAIALYDLKVQK-----NSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLGLP 422
QY 423 QNTWKIGIMDEERTTNLAKACIAAADRVVINTGFLDRTGDEIHTSMGPAEVAFTCELSRVEDVLGLP 476
DB 417 PNTLKGIMDEERTTNLAKACIAAADRVVINTGFLDRTGDEIHTSMGPAEVAFTCELSRVEDVLGLP 476
QY 483 KSQWILAYEDHNDVAGLAAGFSGRAQVKGKGMVMTMELMADVETKIAQPRAGASTAWVP 542
DB 477 KSTPWIKAYERNVLSGLFCGLRKAQIGKGMWAMPDLMDMYKQGDQLRAGANTAWVP 536
QY 543 SPTAATLHALHYHQDVAAVVOQGLA-----GKRATTIEQLITPLAKELAWADEIREEV 597
DB 537 SPTAATLHALHYHQDVAAVVOQGLA-----GKRATTIEQLITPLAKELAWADEIREEV 597
QY 598 DNNQOSILGYVVRVVOQGVGSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITSADV 657
DB 597 DNNVQILGYVVRVVEQIGGSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITSADV 656
QY 716 RASLERMAPLVDRQVAGVAVRPMAPNPDSDIAFLAQELILSGAQOQNGYTPILHR-R 716
DB 657 QASLENNAKVVDQVQVAGDPAYRPMAGNFANSCAFKAASDLIFLGVKQNGYTPILHAWR 716
QY 717 RRE 719
DB 717 LRE 719

RESULT 12
QBPOR6
ID QBPOR6 PRELIMINARY; PRT; 538 AA.
AC QBPOR6
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Malate synthase.
GN MLS OR XAC0256.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP "SEQUENCE FROM N.A.

RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral J.C., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Forghieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lenos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Medeiros J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spicola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
DR EMBL; AB011650; AAM35148.1;
DR InterPro; IPR006252; Malate_synthase.
DR InterPro; IPR001465; Malate_synthase.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01344; malate_syn_A; 1.
KW Complete proteome.
SQ SEQUENCE 538 AA; 59271 MW; 80A47F8D32B7C848 CRC64;

Query Match 7.3%; Score 279.5; DB 16; Length 538;
Best Local Similarity 24.9%; Pred. No. 9.2e-10;
Matches 123; Conservative 59; Mismatches 204; Indels 107; Gaps 20;
QY 249 GTTDRAGVKDVILSAITTIMDFEDSVAAVDAADKVLGYRNWLGKGLAAVVDKDTA 308
DB 102 GPTDFKWINALNSCAKVFMAFDFTAPT-----WRNLLA-GQTLVAAV----- 146
QY 309 FLVLNLRNDRNTAPGGGFTL-----PGRSLMFRVNVG-HLMTNDALVDTDGSEVEGIMDA 364
DB 147 -----RGDLSFDAPNGKHYALRPEAEARVLIIVRPRGHLDEKHVLI--DGQFLAGGLFDA 199
QY 365 LFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLGLPQN 424
DB 200 ---ALFAFH-----NGRTLLAKDRGPYLYLPKLSMEEAALMDTALAHTEAMGLPHG 249
QY 425 TWKIGIMDEERTTNLAKACIAAADRVVINTGFLDRTGDEIHTSMGPAEVAFTCELSRVEDVLGLPQN 481
DB 250 QIKVTVLITLPAVPEMDEILYALFERIVGLNCGRWYIFSYLKTFFRAHRRVLPFERQV 309
QY 482 KXSQWILAYED-----HNVDAGLAAGFSGRAQVKGKGMVMTMELMADVETKIAQPRAG 535
DB 310 TWQGLFLRAYSELLIKTCHRRGAHANGVAAQIPINHDAAEQANMARVADKUREVSAG 369
QY 536 ASTAWVPSP-----AATLHALHYHQDVAAVVOQGLAGKRRATIEQLITPLA 583
DB 370 HDGTWVAHPALIPVAMKLFDEHMTAHOQHVLNRNDVQVTR-----DLLIAP-- 415
QY 584 KSLANAPDEI-REEVDNNCOSILGYVVRVVOQGVGSKVPDIHDVVALMEDRATLRISSQL 642
DB 416 -----SPGNVTRAGFEGNVVCVRYLAALWLD-GNGC--VPIHH-----LNEADTAETISRAQ 464
QY 643 LANWLRHGV-----ITSADVRASLERMAPLVDRQVAGVAVRPMAPNPDSDIAFLAQELILSGAQOQNGYTPILHR-R 691
DB 465 LQWQLHHGQHLDDGTAIHQHLLQATLRALPARL-----GTATLPGAARIDEAL 515
QY 692 LAAQELILSGAQO 704
DB 516 LEE-----LSGADE 524
RESULT 13
QBPOR6
ID QBPOR6 PRELIMINARY; PRT; 540 AA.
AC QBPOR6

Db 609 EKHVLHNPQTSQS-----LFDGLFVFNHAKA-----LIAQSGP-YFYLEKLQ 652
QY 401 GPAEVAFTCELFSESRVEDVLGLPQNTMKIGIMDEERTTVNLKACIKAAADRVVFEINTGFL 460
Db 653 SAEQAQLWADVFKYTEDKLGARGTITKCTVLIEHLASFOJHEIHALKONIVGLNGCRW 712
QY 461 DRTGDEIHTSME-----AGPMVRKGTMSQPWILAYEDHNDVAGLAAGFSG 506
Db 713 DYIFSYIKTFQNHKFKLLPDRFQIGMTAPFMR-----NYSLEVIKACHLRG 758
QY 507 RAQVCKGMWMTMELMADWV--ETKIAQPR-----GASTAVNPSPPTATLHALHYHQ 556
Db 759 IHAMG-GVAAQIPKHDQVANDKAFALVRADKEREATDGHGTWVHP-----805
QY 557 VDVAVQOGLAGKRRATIEQLLTP--LAKELAWA-----PDEIREVD--NNCQ 602
Db 806 -----GLVPLAKRVFDQMPKPNQISKNLTRANCTKEDLTVIPEGTRTEAGFRHNS 857
QY 603 SILGYVVRVWDQGVGCSKVPDIHDVLMEDRATLRISQLLANWLH-----G 650
Db 858 VTLGYLDSWL-RGTGC--VPLYN---LMEDAATAEISRAQLWQWLHHDKLEDRGTIDAG 911
QY 651 VI---TSADVRSASLERMAPLYDR 670
Db 912 LVKQTIAAETERRLLRAGSVVNR 934

Search completed: November 21, 2003, 16:08:06
Job time : 40.4127 secs

RESULT 15
ID 017353 PRELIMINARY; PRT; 968 AA.
AC 017353;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 108.6 kDa protein.
GN C05E4.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Blanchard M., Kramer J., Gibson A.,
RA "The sequence of C. elegans cosmid C05E4.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RA "Direct Submission";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF026209; AAB71278.2; -;
DR HSSP; O53752; 1F61.
DR WormPep; C05E4.9; CE23521.
DR InterPro; IPR006254; Isocit lyase.
DR InterPro; IPR000918; Isocit lyase ph.
DR InterPro; IPR006252; Malate synthase.
DR InterPro; IPR001465; Malate synthase.
DR Pfam; PF00463; ICL; 1.
DR Pfam; PF01274; Malate synthase; 1.
DR ProDom; PD001857; Isocit lyase ph; 1.
DR TIGRFAMs; TIGR01346; isocit lyase; 1.
DR TIGRFAMs; TIGR01344; malate syn A; 1.
DR PROSITE; PS00161; ISOCITRATE_LYASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 968 AA; 108628 MW; 7BAAF979C1F6D31F CRC64;

Query Match 5.2%; Score 198.5; DB 5; Length 968;
Best Local Similarity 23.8%; Pred. No. 0.00044;
Matches 148; Conservative 70; Mismatches 222; Indels 183; Gaps 34;

QY 127 LNAANARNGSLYDALYGVDPETDGAEGKTYNKGDKVIAY--ARKELDSVPLSSG 184
Db 416 LSSITALSSTEEAQFQFAVASQ--DEELSLTAQNAGVAGDEKILTPDARFLHDL----- 458
QY 185 SFGDATGTVDGQIV-----VALPD--KSTGLANPQFAGYTGAESPTSV 229
Db 469 ---NTEPNRRLRLSKENQVQADINNSLWFPDFNKETEVLRSQ--GWKG-AEIPDL 521
QY 230 LLINHLHILIDPESQVGTTRAGVKDVLESATITIMDFEDSVAAVDAADKVLGYRN 289
Db 522 ---QDRVEI-----TGPTDRWVNAVNSGANVFMAFDENSPST-----WRN 562
QY 290 WL--GLNKGDLAAAVDKGTAFRLVNRDRNYTAP--GGQFTLPGR--SLMFRNVG--HLM 344
Db 563 QLEGQINLYD-----AVRNINISYTHPTTKKEYTLNEKHAVLKVRPGWHL 608
QY 345 TNDALV---DTDGEVFEIGINDALFT--GLIAHGLKASDVNGPLINSGTISYIVKPKMH 400

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 25.6707 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-4

Perfect score: 2193

Sequence: 1 VQYGGSSVADAEIRIRVAE.....SAATRRPCRTGRDGRWACQ 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq 19Jun03.*
1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2193	100.0	1172	22 AAU01900	M. tuberculosis Tb
2	1619.5	73.8	421	22 AAU08233	Mycobacterium poly
3	1248.5	56.9	421	22 AAG64046	Corynebacterium th
4	1227.5	56.0	421	21 AAB29607	Modified Corynebac
5	1227.5	56.0	421	24 ABP97764	Amino acid sequenc
6	1224.5	55.8	420	24 ABP97763	Amino acid sequenc
7	1224.5	55.8	421	22 AAU71874	C. glutamicum meta
8	1224.5	55.8	421	22 AAG90027	C. glutamicum prote
9	1224.5	55.8	421	22 AAB79646	Corynebacterium gl

10	1223.5	55.8	421	15 AAR51469	AK alpha subunit T
11	1223.5	55.8	421	15 AAR51465	AK alpha subunit T
12	1223.5	55.8	421	16 AAR74416	C. glutamicum prot
13	1223.5	55.8	421	16 AAR86307	Corynebacterium gl
14	1223.5	55.8	421	19 AAW68147	Aspartokinase alph
15	1223.5	55.8	421	19 AAW69547	Bravibacterium lac
16	1223.5	55.8	421	19 AAW47401	B. lactofermentum
17	1219.5	55.6	421	14 AAR40186	B. flavum AK Bre
18	1219.5	55.6	421	15 AAR51466	AK wild type alpha
19	1219.5	55.6	421	15 AAR51470	AK wild type alpha
20	1219.5	55.6	421	18 AAW06586	B. lactofermentum
21	1219.5	55.6	421	19 AAW68152	Aspartokinase alph
22	1219.5	55.6	842	15 AAR63567	Feedback inhibito
23	1219.5	55.6	842	15 AAR63570	Feedback inhibito
24	1218.5	55.6	421	15 AAR63571	Feedback inhibito
25	1217.5	55.5	421	22 AAE04857	Corynebacterium gl
26	1217.5	55.5	421	23 ABE81982	C. glutamicum aspa
27	1215.5	55.4	421	15 AAR63566	Feedback inhibito
28	1215.5	55.4	842	15 AAR63568	Feedback inhibito
29	1215.5	55.4	842	15 AAR63569	Feedback inhibito
30	1211.5	55.2	421	22 AAE04867	Corynebacterium sp
31	1211.5	55.2	421	23 ABB81989	C. glutamicum aspa
32	1176	53.6	505	22 AAU54464	Propionibacterium
33	787	35.3	409	21 AAB29516	Methylophilus meth
34	785	35.8	411	14 AAR40846	Aspartokinase II.
35	773.5	35.3	254	23 ABP65555	Bifidobacterium lo
36	770	35.1	156	20 AAV39178	M. tuberculosis an
37	770	35.1	156	20 AAV39035	M. tuberculosis re
38	703	32.1	384	24 ABP79567	N. gonorrhoeae ami
39	703	32.1	404	23 ABE48331	Listeria monocytog
40	673	30.7	394	19 AAW98712	H. pylori GHPO 774
41	612.5	27.9	606	11 AAR06915	Protein with aspar
42	563	25.7	424	23 ABP40380	Staphylococcus epi
43	506	23.1	403	23 ABB47953	Listeria monocytog
44	419	19.1	118	23 ABP08315	Human OREF protein
45	377	17.2	564	22 AAE04359	Glycine max aspart

ALIGNMENTS

RESULT 1
AAU01900
ID AAU01900 standard; Protein; 1172 AA.
XX AAU01900;
AC
XX
DT 29-AUG-2001 (first entry)
XX
DE M. tuberculosis TbF14 fusion protein.
XX
KW TbF14; antigen; vaccine; tuberculosis; AIDS; His tag; MtB81; Mo2;
KW acquired immunodeficiency disease.
XX
OS Synthetic.
OS Mycobacterium tuberculosis.
XX
XX
FH Key Location/Qualifiers
FT Binding-site 3..8 /label= Histidine tag
FT FT /note= "Nickel chelating region used for purifying
FT FT the fusion protein"
FT FT
FT Region 9..749 /label= MtB81_region
FT FT
FT Misc-difference 49 /note= "Encoded by TC"
FT FT 750..1172
FT Region /label= Mo2_region
XX
XX WO200124820-A1.
XX
XX 12-APR-2001.
PD
XX

```

PF 10-OCT-2000; 2000WO-US28095.
XX
XX 07-OCT-1999; 99US-0158338.
PR 07-OCT-1999; 99US-0158425.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
XX WPI; 2001-290576/30.
DR N-PSDB; AAS03794.
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens -
XX
XX Claim 3; Fig 3; 168pp; English.
XX
XX The sequence represents Mycobacterium fusion protein antigen TbF14
XX consisting of a His tag for purification, antigen McB81 and antigen MO2.
CC Compositions comprising at least 2 heterologous antigens, as a fusion
CC protein, and vectors expressing the fusion proteins are used as vaccines
CC to prophylactically immunise mammals (especially humans) against
CC infection by Mycobacteria. The compositions contain at least 2
CC heterologous antigens that increase the serological sensitivity of
CC individuals infected with tuberculosis, a disease frequently affecting
CC patients with acquired immunodeficiency disease, AIDS.
XX
XX SQ Sequence 1172 AA;
XX
XX Query Match 100.0%; Score 2193; DB 22; Length 1172;
XX Best Local Similarity 100.0%; Pred. No. 5.5e-185;
XX Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 VKYGGSSVADAEIRRVAAERIVATKQGNVTVVVSAMGDTTDDLDAQVCPAPP 60
XX 749 VKYGGSSVADAEIRRVAAERIVATKQGNVTVVVSAMGDTTDDLDAQVCPAPP 808
XX
XX 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIVTPGRLOTA 120
XX 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIVTPGRLOTA 868
XX
XX 121 LEEGRVVLVAGFGVQSODTKDVTTLGRGSGDITAVAMAALGADVCETVDVDFSSADP 180
XX 869 LEEGRVVLVAGFGVQSODTKDVTTLGRGSGDITAVAMAALGADVCETVDVDFSSADP 928
XX
XX 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
XX 929 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 988
XX
XX 241 IKDVPMEDEPILTCVAHDRSEAKVTIVGLPDIPGYAAKVFAVARRRORHGAERLOGR 300
XX 989 IKDVPMEDEPILTCVAHDRSEAKVTIVGLPDIPGYAAKVFAVARRRORHGAERLOGR 1048
XX
XX 301 GRQDRHLLHPQTSGPPPPKNTSRSETRSASTQLLYDDHIGKVLIGAGMRSHPGVTAT 360
XX 1049 GRQDRHLLHPQTSGPPPPKNTSRSETRSASTQLLYDDHIGKVLIGAGMRSHPGVTAT 1108
XX
XX 361 FCEALAAVGNIELISTSDQSRCCAAATPNWTRPWSRCMKESGSAATRRPCTGRDGR 420
XX 1109 FCEALAAVGNIELISTSDQSRCCAAATPNWTRPWSRCMKESGSAATRRPCTGRDGR 1168
XX
XX 421 WACQ 424
XX 1169 WACQ 1172
XX
XX RESULT 2
XX AAU08233
XX AC AAU08233;
XX AC AAU08233;
XX DT 17-DEC-2001 (first entry)

```

```

XX Mycobacterium polypeptide for MO-2.
XX
XX Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
XX immunostimulant; MO-2.
XX
XX Mycobacterium sp.
XX
XX Key Location/Qualifiers
XX Misc-difference 41
XX Misc-difference /note= "Encoded by T"
XX Misc-difference 305
XX Misc-difference /note= "Encoded by GAC"
XX
XX WO200162893-A2.
XX
XX 30-AUG-2001.
XX
XX 26-FEB-2001; 2001WO-US05992.
XX
XX 25-FEB-2000; 2000US-0185037.
XX 08-AUG-2000; 2000US-0223828.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Skeiky Y, Owendale P, Jen S, Lodes M;
XX WPI; 2001-536638/59.
XX N-PSDB; AAS12505.
XX
XX An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
XX Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
XX response to and inhibiting development of a Mycobacterium infection -
XX Claim 1; Fig 17; 161pp; English.
XX
XX The present invention relates to the isolation of Mycobacterium
XX tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
XX encoding them. The invention describes compounds and methods for the
XX diagnosis of tuberculosis or for inducing protective immunity against
XX tuberculosis. The compounds comprise at least one immunogenic portion
XX of one or more Mycobacterium proteins and nucleic acid molecules
XX encoding such polypeptides. The Mycobacterium proteins and nucleic acid
XX molecules encoding them can be used in diagnostic kits for the detection
XX of Mycobacterium infection in patients and biological samples. The
XX compounds of the invention and antibodies directed against the
XX Mycobacterium proteins may be used in vaccines for immunisation against
XX Mycobacterium infections. The nucleic acids encoding the Mycobacterium
XX proteins may be used in gene therapy. The present sequence represents
XX Mycobacterium polypeptide for MO-2.
XX
XX SQ Sequence 421 AA;
XX
XX Query Match 73.8%; Score 1619.5; DB 22; Length 421;
XX Best Local Similarity 85.1%; Pred. No. 3.8e-142;
XX Matches 338; Conservative 8; Mismatches 34; Indels 17; Gaps 3;
XX
XX 1 VKYGGSSVADAEIRRVAAERIVATKQGNVTVVVSAMGDTTDDLDAQVCPAPP 60
XX 5 VKYGGSSVADAEIRRVAAERIVATKQGNVTVVVSAMGDTTDDLDAQVCPAPP 64
XX
XX 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIVTPGRLOTA 120
XX 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIVTPGRLOTA 124
XX
XX 121 LEEGRVVLVAGFGVQSODTKDVTTLGRGSGDITAVAMAALGADVCETVDVDFSSADP 180
XX 125 LEEGRVVLVAGFGVQSODTKDVTTLGRGSGDITAVAMAALGADVCETVDVDFSSADP 184
XX
XX 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
XX 185 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 244

```

Qy	241	IKDVPNEDPILTCGVAHDSSEAKVTIVGLPDIPGYAAKVFRVARRRRQHRHGAAE	-----	299
Db	245	IKDVPNEDPILTCGVAHDSSEAKVTIVGLPDIPGYAAKVFRVARRRRQHRHGAAE	-----	295
Qy	296	RLQGRGRQRDRHLLHLLPQTS---GPPPWKNWTRSETRSASTQLLYDDHIGKVSLLIGAGMR	352	
Db	296	VLQNVSKVEKGKTDITFTCSRVDGPAAVEKLDLSRNEIGFSQLLYDDHIGKVSLLIGAGMR	355	
Qy	353	SHPGVTATFCEALAAVGVNIELISTSEDRSRCCAAT	389	
Db	356	SHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDT	392	
RESULT 3				
AAG64046				
ID	AAG64046 standard; Protein; 421 AA.			
XX	AAG64046;			
DT	11-SEP-2001 (first entry)			
DE	Corynebacterium thermoaminogenes lysin biosynthetic enzyme lysC.			
KW	Heat-resistant; lysin biosynthesis; enzyme; coryneform;			
KW	aspartate-semialdehyde dehydrogenase; lysC.			
XX	Corynebacterium thermoaminogenes.			
XX	JP2001120270-A.			
PN	08-MAY-2001.			
PD	01-NOV-1999; 99JP-0311148.			
PF	01-NOV-1999; 99JP-0311148.			
PR	(AJIN) AJINOMOTO KK.			
PA	WPI; 2001-364760/38.			
XX	N-PSDB; AAH45374.			
DR	A heat-resistant lysin biosynthetic system enzyme gene of a high			
XX	temperature-resistant coryneform microbe -			
PT	Claim 6; Page 21-22; 27pp; Japanese.			
XX	The invention relates to a gene from a high temperature-resistant			
CC	coryneform microbe that encodes a heat-resistant lysin biosynthetic			
CC	enzyme. The enzyme has aspartate-semialdehyde dehydrogenase activity			
CC	and can be used for growing amino acid-producing microbes. The			
CC	present amino acid sequence corresponds to an enzyme of the			
XX	invention.			
XX	Sequence 421 AA;			
Qy	Query Match 56.9%; Score 1248.5; DB 22; Length 421;			
Db	Best Local Similarity 65.2%; Pred. No. 1.5e-107;			
XX	Matches 251; Conservative 58; Mismatches 63; Indels 13; Gaps 3			
Qy	1	VQYGGSSVADAERIRRVAEIRVATKQGNDDVVVWSAMGDTTDDLDDLAQQVCPPPPR	60	
Db	5	VQYGGSSLESARIRNVAEIRVATKAGNDVVVVCAMGDTTDELLDLAAANVPVPPAR	64	
Qy	61	ELDMLLTAGERISNALVAMATESIGAHARSTGQAGVITGTGHNAKIIDVTPGELQTA	120	
Db	65	EMDMLLTAGERISNALVAMATESIGAEAQSTGQAGVLTTERHGNARIVDVTPGVRREA	124	
Qy	121	LEEGRVVLVAGFGQYSQDTKVITLLGRGSGDTTAVAMAAALGADVCEIYTDVDGIESADP	180	
Db	125	LEEGKICIVAGFGQGNKETRVITLLGRGSGDTTAVALLAALGADVCEIYSDVDGVITADP	184	
Qy	181	RIVRNARKLDTVTFFEMLEMAACGAKVLMRLRCVEYARHNI PVHVRSSYSDFGTVVVGS	240	

Db	185	RIVPNAQKLERUSFEMLEAAVSGSKILVRSVEYAFNVPMKVRSSYSNDPCTLIAGS	241
Qy	241	IKQVPMEDPILTVGAHDSRSEAKVTIVGLPDIPGVAAKVFRAVARR-----RQHRHGAA	294
Db	245	MEDIPMEEAVLTVGATDKSEAKVTIVGLIPDKPGEAAKVFRAALADAEINIDMWLQNVSSVE	304
Qy	295	ERLQGRGRODRHHILLPQTGSPPPKWNWTFSETRSASTOLLVDDDHIGKVSILGAGNRSH	354
Db	305	D-----GTTD--ITTCPSRSDGRAMELLKKMQQGGDWNTLVDDQVGRKVSILGAGMKSH	357
Qy	355	PGVTATFCEALAAVGVNIELISTSE	379
Db	358	PGVTAEFMEALRDVNVNVELISTSE	382
RESULT 4			
AA	AB29607		
ID	AB29607	standard; Protein; 421 AA.	
XX	AA		
XX	AA	AB29607;	
XX	DT	19-FEB-2001 (first entry)	
XX	DE	Modified Corynebacterium Glutamicum aspartokinase, T3111.	
XX	KW	Aspartokinase; AK; lysine synthesis; feedback inhibition;	
KW	KW	L-lysine; L-threonine; insensitive; product inhibition;	
KW	KW	coryneform bacterium; mutant; mutain.	
OS	OS	Corynebacterium glutamicum.	
OS	OS	Synthetic.	
XX	XX	Key	Location/Qualifiers
FT	FT	Misc-difference 311	/note= "Wild-type Thr replaced with Ile"
XX	XX	WO200063388-A1.	
XX	XX	26-OCT-2000.	
XX	XX	14-APR-2000; 2000WO-JP02456.	
XX	XX	19-APR-1999; 99JP-0110437.	
XX	PA	(KYOW) KYOWA HAKKO KOGYO KK.	
XX	PI	Yokoi H, Ohnishi J, Ochiai K, Yonetani Y, Ozaki A;	
XX	DR	WPI; 2000-687179/57.	
XX	DR	N-PSDB; AAC64119.	
PT	PT	DNA encoding modified aspartokinase without synergistic feedback	
PT	PT	inhibition by L-lysine and L-threonine for efficient production of	
XX	XX	L-lysine by coryneform bacterium fermentation -	
XX	PS	Claim 1; Page 29-30; 33pp; Japanese.	
XX	CC	The invention relates to a modified aspartokinase from Corynebacterium	
CC	CC	glutamicum (e.g., AB29607) and to DNA encoding the modified	
CC	CC	aspartokinase (e.g., AAC64119). The aspartokinase of the invention is	
CC	CC	modified to remove its sensitivity to feedback inhibition by L-threonine	
CC	CC	or L-lysine. This is achieved by substituting the wild-type Thr 311 with	
CC	CC	any other amino acid; for example, in AB29607 Thr 311 is substituted by	
CC	CC	Ile. The invention also relates to expression vectors and host	
CC	CC	coryneform bacteria containing the modified DNA, and to the preparation	
CC	CC	of L-lysine using coryneform bacteria transformed with the modified	
CC	CC	aspartokinase-encoding DNA. Culturing coryneform bacteria transformed	
CC	CC	with the modified aspartokinase DNA of the invention provides efficient	
CC	CC	production of L-lysine. The present sequence represents a modified	
XX	XX	Corynebacterium glutamicum aspartokinase of the invention, T3111.	
XX	XX	Sequence 421 AA;	
XX	XX	Sequence 421 AA;	

Query Match 56.0%; Score 1227.5; DB 21; Length 421;
 Best Local Similarity 64.0%; Pred. No. 1.3e-105;
 Matches 251; Conservative 56; Mismatches 58; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPP 60
 DB 5 VKYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELELAAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAVAIESLGAHARSFTGSOAGVITTCGHNAKIIDVTGRLQTA 120
 DB 65 EMDMLLTAGERISNALVAVAIESLGAHARSFTGSOAGVITTCGHNARIVDVTGVRVREA 124

QY 121 LEEGRVVLVAGFGVSGDVKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGSADP 180
 DB 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184

QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSRPGTVVVG 240
 DB 185 RIVPNAQKLEKLSFEEMLELAAGVSKILVLRSEYARAFNPLRVSRSSYNDPGLIAGS 244

QY 241 IKDVPMDPILTCVAHDSREAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
 DB 245 MEDIPVEEAULTGVATDKSEAKVTIVLGISDKFGEAAKVFRAADAENIDMVLQNVSSVE 304

QY 291 HGAERLQGRQD-RHHLHLLP--QTSGPPPMKNWTRSETRSASTOLLYDDHIGKVSII 347
 DB 305 DGTDTDIIFTCPRSDGRRAMEILKQLVQG-----NW-----TNVLYDDQVGVKSLV 350

QY 348 GAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 5

ABP97764
 ID ABP97764 standard; Protein; 421 AA.
 AC ABP97764;
 XX
 XX
 DT 28-MAY-2003 (first entry)
 XX
 DE Amino acid sequence of an aspartate kinase carrying T3111 mutation.
 XX
 KW Coryneform bacteria; chemical compound; L-amino acid; vitamin; lysC;
 XX aspartate kinase.
 XX
 OS Synthetic.
 OS Corynebacterium glutamicum.
 PH Key Location/Qualifiers
 FT Misc-difference 1 /note= "Met encoded by GTG"
 FT FT
 FT Misc-difference 311 /note= "wild type Thr changed to Ile"
 FT FT
 XX WO2003014330-A2.
 XX
 PD 20-FEB-2003.
 XX
 XX 30-JUL-2002; 2002WO-EP08465.
 XX
 PR 06-AUG-2001; 2001US-309877P.
 XX
 XX (DECS) DEGUSSA AG.
 XX
 XX Bathe B, Kreutzer C, Moeckel B, Thierbach G;
 XX WPI; 2003-278476/27.
 XX DR N-PSDB; AB277392.
 XX
 XX Coryneform bacteria to produce chemical compounds e.g. L-lysine, has at
 PT least two copies of open reading frame, gene or allele in question at a
 PT particular desired site

XX
 PS Claim 16; Page 105-106; 109pp; English.
 XX
 CC The specification describes Coryneform bacteria which produce chemical
 CC compounds, which instead of a single copy of an open reading frame (ORF),
 CC gene or allele naturally present at a desired locus have at least two
 CC copies of the ORF. Gene or allele, preferably in tandem arrangement.
 CC The modified bacteria are is useful for producing chemical compounds
 CC including L-amino acids, vitamins, nucleosides and nucleotides. The
 CC present sequence is encoded by a lysC gene, and represents an aspartate
 CC kinase carrying the T3111 mutation. Tandem duplications of the lysC
 CC gene are used to construct the bacteria of the invention.
 XX
 SQ Sequence 421 AA;

Query Match 56.0%; Score 1227.5; DB 24; Length 421;
 Best Local Similarity 64.0%; Pred. No. 1.3e-105;
 Matches 251; Conservative 56; Mismatches 58; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPP 60
 DB 5 VKYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELELAAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAVAIESLGAHARSFTGSOAGVITTCGHNAKIIDVTGRLQTA 120
 DB 65 EMDMLLTAGERISNALVAVAIESLGAHARSFTGSOAGVITTCGHNARIVDVTGVRVREA 124

QY 121 LEEGRVVLVAGFGVSGDVKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGSADP 180
 DB 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184

QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSRPGTVVVG 240
 DB 185 RIVPNAQKLEKLSFEEMLELAAGVSKILVLRSEYARAFNPLRVSRSSYNDPGLIAGS 244

QY 241 IKDVPMDPILTCVAHDSREAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
 DB 245 MEDIPVEEAULTGVATDKSEAKVTIVLGISDKFGEAAKVFRAADAENIDMVLQNVSSVE 304

QY 291 HGAERLQGRQD-RHHLHLLP--QTSGPPPMKNWTRSETRSASTOLLYDDHIGKVSII 347
 DB 305 DGTDTDIIFTCPRSDGRRAMEILKQLVQG-----NW-----TNVLYDDQVGVKSLV 350

QY 348 GAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 6

ABP97763
 ID ABP97763 standard; Protein; 420 AA.
 AC ABP97763;
 XX
 XX
 DT 28-MAY-2003 (first entry)
 XX
 DE Amino acid sequence of a wild-type aspartate kinase.
 XX
 KW Coryneform bacteria; chemical compound; L-amino acid; vitamin; lysC;
 XX aspartate kinase.
 XX
 OS Corynebacterium glutamicum.
 PH Key Location/Qualifiers
 FT Misc-difference 1 /note= "Met encoded by ATG"
 FT FT
 FT WO2003014330-A2.
 XX
 PD 20-FEB-2003.
 XX
 XX 30-JUL-2002; 2002WO-EP08465.
 XX
 PT

```

PR 06-AUG-2001; 2001US-309877P.
XX (DEGS ) DEGUSSA AG.
XX
XX Bathe B, Kreutzer C, Moeckel B, Thierbach G;
XX
XX WPI; 2003-278476/27.
XX N-PSDB; ABZ77391.
XX
XX Coryneform bacteria to produce chemical compounds e.g. L-lysine, has at
XX least two copies of open reading frame, gene or allele in question at a
XX Particular desired site
XX
XX Claim 15; Page 102-103; 109pp; English.
XX
XX The specification describes Coryneform bacteria which produce chemical
XX compounds, which instead of a single copy of an open reading frame (ORF),
XX gene or allele naturally present at a desired locus have at least two
XX copies of the ORF, gene or allele, preferably in tandem arrangement.
XX The modified bacteria are is useful for producing chemical compounds
XX including L-amino acids, vitamins, nucleosides and nucleotides. The
XX present sequence is encoded by a wild type lysC gene, and represents
XX aspartate kinase. Tandem duplications of the lysC gene are used to
XX construct the bacteria of the invention.
XX
XX Sequence 420 AA;
XX
XX Query Match 55.8%; Score 1224.5; DB 24; Length 420;
XX Best Local Similarity 64.0%; Pred. No. 2.5e-105;
XX Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
XX
XX 1 VQYGGSSVADAERIRVAERIVATKQGNVYVVSAMGDTTDDLLDLAQVCPAPP 60
XX 4 VQYGGSSLESABERIRNVAERIVATKAGNDVYVVSAMGDTTDELELAANVPVPP 63
XX
XX 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNAKIIDVTPEGRQTA 120
XX 64 EMDMLLTAGERISNALVAMAIESLGAEOFTGSOAGVLTTERHGNARIVDVTPEGRVREA 123
XX
XX 121 LEEGRVVLVAGFGVQSDTKVDTLGRGSDTTAVAMAALGADVCEIYTDVDFISADP 180
XX 124 LDEGKICIVAGFGVQVKNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYADP 183
XX
XX 181 RIVNARKLDTVTTEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 240
XX 184 RIVPNAQKLEKISFEEMLELAAGSKILVRSVEYARAFNVLPRVRSYSDNPDTLIAGS 243
XX
XX 241 IKDVPMDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
XX 244 MEDIPVEEAVLTGVATDKSEAKVTVLGSDKFGAAKVFRALADAEINIDMWLQNVSSVE 303
XX
XX 291 HGAERLQGRQD-RHLLHLLP--QTSQPPPKWTRSETRSASTOLLDDHIGKVSLLI 347
XX 304 DGTDTITFTCPRSDDGRRAMEILKKLQVQG-----TNVLYDDQVGVKVSLLV 349
XX
XX 348 GAGKSHPGVGTATFCEALAAVGNIELISTSE 379
XX 350 GAGKSHPGVGTAEFMEALRDVNVNIELISTSE 381
XX
XX RESULT 7
XX AAU71874
XX ID AAU71874 standard; Protein; 421 AA.
XX
XX AC AAU71874;
XX
XX DT 26-FEB-2002 (first entry)
XX
XX DE C. glutamicum metabolic pathway protein encoded by gene #9.
XX
XX KW Metabolic pathway protein; MP; lysine biosynthesis pathway;
XX methionine biosynthesis pathway; large-scale production of fine chemical;
XX Corynebacterium diptheriae; diptheria.

```

```

XX OS Corynebacterium glutamicum.
XX PN WO200166573-A2.
XX
XX PD 13-SEP-2001.
XX
XX PF 22-DEC-2000; 2000WO-IB02035.
XX
XX PR 09-MAR-2000; 2000US-187970P.
XX PR 23-JUN-2000; 2000US-0606740.
XX
XX PA (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;
XX Lee H, Hwang B;
XX
XX WPI; 2001-582269/65.
XX N-PSDB; AAS96084.
XX
XX Nucleic acids encoding metabolic pathway proteins from Corynebacterium
XX glutamicum, useful for producing methionine and lysine in
XX Corynebacterium and Brevibacterium -
XX
XX Disclosure; Page 175-176; 316pp; English.
XX
XX The present invention relates to the isolation of novel Corynebacterium
XX glutamicum genes (AAS96073-AAS96132) encoding metabolic pathway (MP)
XX proteins. The metabolic pathway proteins of the invention
XX include enzymes involved in the lysine and methionine biosynthetic
XX pathways, the polynucleotide sequences of the invention can be used
XX for the large-scale production and/or modulation of expression of
XX fine chemicals such as lysine and methionine. The sequences of the
XX invention may be used to identify C. glutamicum and related organisms
XX e.g. C. diptheriae in a subject to detect diptheria.
XX AAU71863-AAU71922 represent the novel C. glutamicum metabolic pathway
XX proteins of the invention.
XX
XX Sequence 421 AA;
XX
XX Query Match 55.8%; Score 1224.5; DB 22; Length 421;
XX Best Local Similarity 64.0%; Pred. No. 2.5e-105;
XX Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
XX
XX 1 VQYGGSSVADAERIRVAERIVATKQGNVYVVSAMGDTTDDLLDLAQVCPAPP 60
XX 5 VQYGGSSLESABERIRNVAERIVATKAGNDVYVVSAMGDTTDELELAANVPVPP 64
XX
XX 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNAKIIDVTPEGRQTA 120
XX 65 EMDMLLTAGERISNALVAMAIESLGAEOFTGSOAGVLTTERHGNARIVDVTPEGRVREA 124
XX
XX 121 LEEGRVVLVAGFGVQSDTKVDTLGRGSDTTAVAMAALGADVCEIYTDVDFISADP 180
XX 125 LDEGKICIVAGFGVQVKNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYADP 184
XX
XX 181 RIVNARKLDTVTTEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 240
XX 185 RIVPNAQKLEKISFEEMLELAAGSKILVRSVEYARAFNVLPRVRSYSDNPDTLIAGS 244
XX
XX 241 IKDVPMDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
XX 245 MEDIPVEEAVLTGVATDKSEAKVTVLGSDKFGAAKVFRALADAEINIDMWLQNVSSVE 304
XX
XX 291 HGAERLQGRQD-RHLLHLLP--QTSQPPPKWTRSETRSASTOLLDDHIGKVSLLI 347
XX 305 DGTDTITFTCPRSDDGRRAMEILKKLQVQG-----TNVLYDDQVGVKVSLLV 350
XX
XX 348 GAGKSHPGVGTATFCEALAAVGNIELISTSE 379
XX 351 GAGKSHPGVGTAEFMEALRDVNVNIELISTSE 382
XX

```

```

RESULT 8
AAG90027
ID AAG90027 standard; Protein; 421 AA.
XX
AC AAG90027;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 3781.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
DR N-PSDB; AAH65246.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17; SEQ ID NO: 3781; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from Coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 421 AA;
Query Match 55.8%; Score 1224.5; DB 22; Length 421;
Best Local Similarity 64.0%; Pred. No. 2.5e-105;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VKYGGSSVADAEIRRVARIVATKQGNDDVVVVVVSAMGDTDDLDAQQVCPAPPPR 60
DB 5 VKYGGSSLESABIRNVAERIVATKAGNDVVVVCAMGDTTDELLEAAAVNPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAISSLGNHARSFTSGAGVITTTGTHGNAKIIVTFCRLTA 120
DB 65 EMDMLLTAGERISNALVAMAISSLGAOSFTSGAGVLTTERHGNARIVDTPGRVREA 124
QY 121 LEEGRVVLVAGFGQYSQDTKDYTTILGRGSDTAVAMAAALGADVCEIYTDVGDIFSDAP 180
DB 125 LDEGKICIVAGFGQYKNERDVTTLGRGSDTAVALLAALNADVCYISYDVGVTADP 184
QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRLCVRHNPVHVRSSYSRDPGTVVVG 240

```

```

Db 185 RIVPNAQKLEKLSFEEMLEAAVSGSKILVLRSEYARAFNVPLRVSSYSNDEGILLIAGS 244
QY 241 IKDVFEMEDPILTGVAHDRSEAKVTIIVGLPDIPIGYAAKVFRAVA-----RRRRQHR 290
Db 245 MEDIPVEEAVLTGVTADKSEAKVTILGISDKPGEAAKVFRAALADABINIDMWLQNVSSVE 304
QY 291 HGAERLQGRGROD-RHLLHLIP--QTSGPPPKNKNWTRSETRSASTQLLYDDHIGKVSLLI 347
Db 305 DGTDTITFTCPRSDDRAMEILKKLQVQG-----TNVLYDDQVGKVSLLV 350
QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
Db 351 GAGWKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 9
AAB79646
ID AAB79646 standard; Protein; 421 AA.
XX
AC AAB79646;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:26.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100843-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00923.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99DE-1032232.
PR 14-JUL-1999; 99DE-1032922.

```


PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
PA (BADI) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-137957/14.
DR N-PSDB; AAF71765.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
PS Claim 20; Page 193-194; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (NP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
SQ Sequence 421 AA;
Query Match 55.8%; Score 1224.5; DB 22; Length 421;
Best Local Similarity 64.0%; Pred. No. 2.5e-105;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADAEIRIRVAERIVATKQGNDDVVVVVSAAGDTTDDLLDAQVCPAPP 60
DB 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVVCSAGDTTDELELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNAKIIVTPGRLOTA 120
DB 65 EMDMLLTAGERISNALVAMAEISLGAEAQSFSGQAGVLTTERHGNARIVDVTGVRVREA 124
QY 121 LEEGRVVLVAGFOGVQSDTKVDTTLGRGSDTTAVAMAAALGADVCEIYTDVGDGIFSADP 180
DB 125 LDEGKICIVAGFOGVNKTETRDVTTLGRGSDTTAVALLAALNADVCEIYSDVDGVYTADP 184
QY 181 RIVRNARKLDTVTPEMLEMAACGAKVLMRLCIVEYAREHNI PVHVRSSYSYSDRPGTVVVG 240
DB 185 RIVPNAQKLEKLSPEMLELAAGSKILVLSVEYARAFNVPFLVRSYSYNDPOTLIAGS 244
QY 241 IKDVPMEPIITGVADHDESEAKVTIVGLPDIPGVAANKVFAVA-----RRRQHR 290
DB 245 MEDIPVEEAVLTGVDKSEAKVTIVGLISDKPGEAAKVFRLADAEINIDMVLQNVSSVE 304

QY 291 HGBAERLQGRQD-RHHLLLP--OTSPPPKWTRSETRTSASTOLLVDDHIGKVSLLI 347
DB 305 DGTDTITFTCPRSDDGRRAMEILKKQVQG-----TNVLYDDQVGKSLV 350
QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382
RESULT 10
AARS1469
ID AARS1469 standard; Protein; 421 AA.
XX AARS1469;
XX AC AARS1469;
XX DT 15-NOV-1994 (first entry)
XX DE AK alpha subunit T279A.
XX KW Mutant; C. glutamicum; aspartokinase; AK; alpha; beta; subunit;
KW substitution; non-acidic; amino acid; Ala; additional; enzyme;
KW feedback inhibition; synergic inhibition; heat stability.
XX OS Corynebacterium glutamicum.
XX
FH Key Location/Qualifiers
FT Misc-difference 279 /label= T279A
FT
XX JP06062866-A.
XX PD 08-MAR-1994.
XX PF 27-APR-1993; 93JP-0101450.
XX PR 28-APR-1992; 92JP-0110292.
XX PA (AJIN) AJINOMOTO KK.
XX WPI; 1994-121127/15.
DR N-PSDB; AAQ61595.
XX
PT The mutant aspartokinase gene - used to produce AK which is
PT removed from feedback inhibition
XX
PS Disclosure; Page 14-16; 28pp; Japanese.
XX
CC This sequence represents a mutant version of the C. glutamicum asparto-
CC kinase (AK) alpha subunit. The mutation is at position 279 and is a
CC corresponds to the substitution of a non-acidic amino acid, pref.
CC Ala. This mutation causes removal of the enzyme from additional
CC feedback inhibition. Inhibition by Thr is removed completely and
CC the synergic inhibition of Lys and Thr is also removed. The new AK
CC has increased heat stability.
XX
SQ Sequence 421 AA;
Query Match 55.8%; Score 1223.5; DB 15; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.1e-105;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADAEIRIRVAERIVATKQGNDDVVVVVSAAGDTTDDLLDAQVCPAPP 60
DB 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVVCSAGDTTDELELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNAKIIVTPGRLOTA 120
DB 65 EMDMLLTAGERISNALVAMAEISLGAEAQSFSGQAGVLTTERHGNARIVDVTGVRVREA 124
QY 121 LEEGRVVLVAGFOGVQSDTKVDTTLGRGSDTTAVAMAAALGADVCEIYTDVGDGIFSADP 180
DB 125 LDEGKICIVAGFOGVNKTETRDVTTLGRGSDTTAVALLAALNADVCEIYSDVDGVYTADP 184

QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSYSDRPGTVVVG 240
 DB 185 RIVPNAQKLEKSFEMLELAAGVKILVRSVEYARAVNPVLRSSYSNDPGTLIAGS 244
 QY 241 IKDVPMDPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
 DB 245 MEDIPVEEAVLTGATDKSEAKVTVLGSDKPGCEAAKVFRAALADAEINIDMVLQNVSSVE 304
 QY 291 HGAERLQGRGROD-RHHHLHP--QTSGPPPPKNTSRSEASSTQLLYDDHIGKVSLLI 347
 DB 305 DGTDTITFTCPRADGRAMEILKLVQVQ-----TNVLYDDQVQKVSLLV 350
 QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 11

AAR51465
 ID AAR51465 standard; Protein; 421 AA.

AC AAR51465;

DT 15-NOV-1994 (first entry)

XX AK alpha subunit T279A from clone ATCC 13869.

XX Mutant; C. glutamicum; aspartokinase; AK; alpha; beta; subunit;
 KW substitution; non-acidic; amino acid; Ala; additional; enzyme;
 KW feedback inhibition; synergic inhibition; heat stability.

XX Corynebacterium glutamicum.

XX Key Location/Qualifiers
 FH Misc-difference 279
 FT /label= T279A

XX JP06062866-A.

XX 08-MAR-1994.

XX 27-APR-1993; 93JP-0101450.

XX 28-APR-1992; 92JP-0110292.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1994-121127/15.

XX N-PSDB; AAQ61599.

XX The mutant aspartokinase gene - used to produce AK which is
 PT removed from feedback inhibition

XX Disclosure; Page 22; 28pp; Japanese.

XX This sequence represents a mutant version of the C. glutamicum asparto-
 CC kinase (AK) alpha subunit. The mutation is at position 279 and
 CC corresponds to the substitution of a non-acidic amino acid, pref.
 CC Ala. This mutation causes removal of the enzyme from additional
 CC feedback inhibition. Inhibition by Thr is removed completely and
 CC the synergic inhibition of Lys and Thr is also removed. The new AK
 CC has increased heat stability.

XX Sequence 421 AA;

XX Query Match 55.8%; Score 1223.5; DB 15; Length 421;
 XX Best Local Similarity 64.0%; Pred. No. 3.le-105;
 XX Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDLDLAQVCPAPPPR 60
 DB 5 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDLLELAAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGVITTTGTHGNAKIIDVTPGRLQTA 120
 DB 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSGVITTTGTHGNAKIIDVTPGRLQTA 124
 QY 121 LEBGRVVLVAGFGVSGODTKDVTTLGRGSDTTAVAVAAALGADVCEIYTDVDFGSADP 180
 DB 125 LDEKICIVAGFGVSGVNETRDVTTLGRGSDTTAVAVAAALGADVCEIYTDVDFGSADP 184
 QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSYSDRPGTVVVG 240
 DB 185 RIVPNAQKLEKSFEMLELAAGVKILVRSVEYARAVNPVLRSSYSNDPGTLIAGS 244
 QY 241 IKDVPMDPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
 DB 245 MEDIPVEEAVLTGATDKSEAKVTVLGSDKPGCEAAKVFRAALADAEINIDMVLQNVSSVE 304
 QY 291 HGAERLQGRGROD-RHHHLHP--QTSGPPPPKNTSRSEASSTQLLYDDHIGKVSLLI 347
 DB 305 DGTDTITFTCPRADGRAMEILKLVQVQ-----TNVLYDDQVQKVSLLV 350
 QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 12

AAR71416
 ID AAR71416 standard; Protein; 421 AA.

AC AAR71416;

DT 18-JAN-1996 (first entry)

XX C. glutamicum protein.

XX Phosphoenolpyruvate-carboxylase; Corynebacterium glutamicum.

XX Corynebacterium glutamicum.

XX WO9506114-A1.

XX 02-MAR-1995.

XX 17-AUG-1994; 94WO-JP01365.

XX 24-AUG-1993; 93JP-0209775.

XX 24-AUG-1993; 93JP-0209776.

XX 05-JUL-1994; 94JP-0153876.

XX (AJIN) AJINOMOTO KK.

XX Izui K, Matsui H, Sugimoto M, Suzuki T;

XX WPI; 1995-106843/14.

XX N-PSDB; AAQ84970.

XX Variant of phospho-enol pyruvate carboxylase - not substantially
 PT inhibited by aspartic acid, is used for efficient production of
 PT amino acids

XX Disclosure; Page 48-51; 77pp; Japanese.

XX The primers given in AAQ84968-69 were used in the PCR amplification
 CC of C. glutamicum ATCC 13869 DNA to isolate the DNA sequences given
 CC in AAQ84970-71.

XX Sequence 421 AA;

XX Query Match 55.8%; Score 1223.5; DB 16; Length 421;
 XX Best Local Similarity 64.0%; Pred. No. 3.le-105;
 XX Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

PT - used in preparation of L-lysine
 PS Claim 2; Page 17-18; 50pp; English.
 XX
 XX This is the wild-type alpha subunit of Brevibacterium lactofermentum
 CC ATCC 13689 (Corynebacterium glutamicum) aspartokinase, an enzyme
 CC involved in the biosynthesis of L-lysine, but which is subject to
 CC feedback inhibition. A claimed recombinant DNA molecule that is
 CC autonomously replicable in coryneform bacteria comprises a DNA
 CC sequence coding for an aspartokinase in which feedback inhibition
 CC by L-lysine and L-threonine is desensitized, and a DNA sequence
 CC coding for a diaminopimelate decarboxylase. This mutant
 CC aspartokinase preferably has residue Ala-279 of the alpha subunit
 CC (see AA68152) and residue Ala-30 of the beta subunit (see AA68153)
 CC replaced by an amino acid that is not Ala and is not an acidic
 CC amino acid. The host coryneform bacteria are used in a claimed
 CC method for producing L-lysine, which is useful as a fodder
 CC additive. They show improvements in L-lysine yield without
 CC suffering restraint of growth.
 XX
 SQ Sequence 421 AA;
 Query Match 55.8%; Score 1223.5; DB 19; Length 421;
 Best Local Similarity 64.0%; Pred. No. 3.1e-105;
 Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
 QY 1 VKYGGSSVADAEIRRVAAERIVATKQGNVWVVSAMGDTTDLDDLAQVCPAPP 60
 Db 5 VKYGGSSLESABERIRNVAERIVATKAGNDVVVCSAMGDTTDLLELAANVPVPP 64
 QY 61 ELDMLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTHGNAKIIDVTPGRLQTA 120
 Db 65 EMDMLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTHGNARIVDVTPGRVREA 124
 QY 121 LEEGRVVLVAGFGVSGDVKDTTLGRGSGDTTAVAMAALGADYCEIYTDVGGIFSADP 180
 Db 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSGDTTAVALAALNADYCEIYSDVDGVITADP 184
 QY 181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRLCVEYARRHNPVHVRSSYSDRPGVVVGS 240
 Db 185 RIVPNAQKLEKLSFEEMLELAAGVKILVLRSEYARAFNPLRVSRSSYSDNPGTLIAGS 244
 QY 241 IKDVMEDPILTCVAHDSREAKVTIVGLPDIGYAAKVERAVA-----RRRQHR 290
 Db 245 MEDIPVEEAVLTGVTADKSEAKVTVLGISDKPGEAAKVFRLADAEINIDVQLQNVSSVE 304
 QY 291 HGAERLQGRGQD-RHHLLHP--QTSPPPPKWNTRSETRSASTQLLYDDHIGKVSLLI 347
 Db 305 DGTDTITFCPRADGRAMEILKQLVQG-----NW-----TNVLYDDQVGKVSIV 350
 QY 348 GAGMRSHPGVTATFCALAAVGNIELISTSE 379
 Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382
 RESULT 15
 AA69547
 ID AA69547 standard; Protein; 421 AA.
 AC
 XX
 XX AA69547;
 DT
 XX
 XX 13-OCT-1998 (first entry)
 DE Brevibacterium lactofermentum lysC protein.
 XX
 XX Brevibacterium lactofermentum; lysC; L-lysine; coryneform bacterium;
 KW aspartokinase; feedback inhibition; dihydrodipicolinate reductase;
 KW diaminopimelate decarboxylase; aspartate aminotransferase.
 XX
 OS Brevibacterium lactofermentum.
 XX
 PN EP854189-A2.
 XX

PD 22-JUL-1998.
 XX
 PF 05-DEC-1997; 97EP-0121443.
 XX
 PR 05-DEC-1996; 96JP-0325659.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 XX
 PI Araki M, Nakamatsu T, Sugimoto M, Yoshihara Y;
 XX WPI; 1998-379060/33.
 DR N-PSDB; AAV40254.
 XX
 XX Recombinant DNA autonomously replicable in coryneform bacteria -
 PT used to produce L-lysine, codes for e.g. aspartokinase,
 PT dihydrodipicolinate reductase and synthase and di-amino-pimelate
 PT decarboxylase
 XX
 PS Claim 2; Page 22-23; 59pp; English.
 XX
 CC The present invention describes a recombinant DNA autonomously
 CC replicable in cells of coryneform bacteria (CB), comprising a DNA
 CC sequence coding for an aspartokinase (AK) in which feedback inhibition
 CC by L-lysine and L-threonine is desensitized, a DNA sequence coding for
 CC a dihydrodipicolinate reductase (DHP), a DNA sequence coding for
 CC diaminopimelate synthase (DHPs), a DNA sequence coding for
 CC diaminopimelate decarboxylase (DAMD) and a DNA sequence coding for
 CC aspartate aminotransferase (AAT). The present sequence is wild-type
 CC lysC from Brevibacterium lactofermentum. The DNA and related products
 CC from the present invention, can be used for improving L-lysine
 CC productivity by CB. The L-lysine produced can be used as a fodder
 CC additive.
 XX
 SQ Sequence 421 AA;
 Query Match 55.8%; Score 1223.5; DB 19; Length 421;
 Best Local Similarity 64.0%; Pred. No. 3.1e-105;
 Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
 QY 1 VKYGGSSVADAEIRRVAAERIVATKQGNVWVVSAMGDTTDLDDLAQVCPAPP 60
 Db 5 VKYGGSSLESABERIRNVAERIVATKAGNDVVVCSAMGDTTDLLELAANVPVPP 64
 QY 61 ELDMLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTHGNAKIIDVTPGRLQTA 120
 Db 65 EMDMLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTHGNARIVDVTPGRVREA 124
 QY 121 LEEGRVVLVAGFGVSGDVKDTTLGRGSGDTTAVAMAALGADYCEIYTDVGGIFSADP 180
 Db 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSGDTTAVALAALNADYCEIYSDVDGVITADP 184
 QY 181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRLCVEYARRHNPVHVRSSYSDRPGVVVGS 240
 Db 185 RIVPNAQKLEKLSFEEMLELAAGVKILVLRSEYARAFNPLRVSRSSYSDNPGTLIAGS 244
 QY 241 IKDVMEDPILTCVAHDSREAKVTIVGLPDIGYAAKVERAVA-----RRRQHR 290
 Db 245 MEDIPVEEAVLTGVTADKSEAKVTVLGISDKPGEAAKVFRLADAEINIDVQLQNVSSVE 304
 QY 291 HGAERLQGRGQD-RHHLLHP--QTSPPPPKWNTRSETRSASTQLLYDDHIGKVSLLI 347
 Db 305 DGTDTITFCPRADGRAMEILKQLVQG-----NW-----TNVLYDDQVGKVSIV 350
 QY 348 GAGMRSHPGVTATFCALAAVGNIELISTSE 379
 Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382
 Search completed: November 21, 2003, 16:03:11
 Job time : 27.6707 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 8.9415 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-4

Perfect score: 2193

Sequence: 1 VKYGGSSVADAEIRRVAAE.....SAATRRPCTRGDRGWAQ 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PGTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1545.5	70.5	421	4	US-08-311-731A-30
2	1223.5	55.8	421	1	US-08-532-828B-3
3	1223.5	55.8	421	1	US-08-700-359-9
4	1223.5	55.8	421	2	US-08-596-366-6
5	1223.5	55.8	421	2	US-08-967-104-6
6	1223.5	55.8	421	3	US-08-985-908-5
7	1223.5	55.8	421	3	US-08-852-730-14
8	1223.5	55.8	421	3	US-08-985-916-5
9	1219.5	55.6	421	1	US-08-532-828B-4
10	823	37.5	461	4	US-05-252-991A-29720
11	785	35.8	411	1	US-07-684-135A-2
12	770	35.1	156	4	US-08-056-356-227
13	770	35.1	156	4	US-09-072-596-222
14	746.5	34.0	438	4	US-08-328-352-6511
15	563	25.7	424	4	US-09-134-001C-5225
16	359.5	16.4	820	2	US-08-380-182-23
17	314.5	14.3	262	4	US-08-887-534A-23
18	314.5	14.3	262	4	US-08-527-431-23
19	313.5	14.3	863	2	US-08-380-182-19
20	310.5	14.2	449	1	US-08-256-136-2
21	310.5	14.2	449	2	US-08-950-737-2
22	310.5	14.2	449	3	US-08-973-461A-8
23	310.5	14.2	449	3	US-08-648-010-8
24	310.5	14.2	863	2	US-08-380-182-20
25	281.5	13.3	463	4	US-08-134-001C-4160
26	285.5	13.0	172	1	US-08-532-828B-5
27	285.5	13.0	172	1	US-08-700-359-11

28	285.5	13.0	172	3	US-08-985-908-7	Sequence 7, Appli
29	285.5	13.0	172	3	US-08-852-730-16	Sequence 16, Appli
30	285.5	13.0	172	3	US-08-985-916-7	Sequence 7, Appli
31	281.5	12.8	172	1	US-08-532-828B-6	Sequence 6, Appli
32	281.5	12.8	172	2	US-08-596-366-8	Sequence 8, Appli
33	281.5	12.8	172	2	US-08-967-104-8	Sequence 8, Appli
34	241	11.0	441	4	US-09-198-452A-1124	Sequence 1124, Ap
35	223	10.2	478	4	US-09-107-532A-6090	Sequence 6090, Ap
36	144	6.6	282	4	US-09-107-532A-6320	Sequence 6320, Ap
37	136	6.2	263	4	US-09-134-001C-4941	Sequence 4941, Ap
38	135	6.2	575	4	US-09-252-991A-22037	Sequence 22037, A
39	127.5	5.8	511	4	US-09-252-991A-28126	Sequence 28126, A
40	124.5	5.7	311	4	US-09-252-991A-21112	Sequence 21112, A
41	124	5.7	248	4	US-09-198-452A-740	Sequence 740, App
42	119	5.4	523	2	US-08-997-080-114	Sequence 114, App
43	119	5.4	523	2	US-08-997-382-114	Sequence 114, App
44	119	5.4	523	3	US-09-095-855-114	Sequence 114, App
45	119	5.4	523	4	US-09-324-542-114	Sequence 114, App

ALIGNMENTS

RESULT 1

US-08-311-731A-30

; Sequence 30, Application US/08311731A

; Patent No. 6583266

; GENERAL INFORMATION:

; APPLICANT: SMITH, DOUGLAS

; APPLICANT: MAO, JEN-I

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 411

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

; STREET: 600 ATLANTIC AVENUE

; CITY: BOSTON

; STATE: MASSACHUSETTS

; COUNTRY: USA

; ZIP: 02210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/311,731A

; FILING DATE:

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: GATES, EDWARD R.

; REGISTRATION NUMBER: 31,616

; REFERENCE/DOCKET NUMBER: C0044/7125

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/720-3500

; TELEFAX: 617/720-2441

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 421 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: MYCOBACTERIUM LEPRAE

US-08-311-731A-30

Query Match 70.5%; Score 1545.5; DB 4; Length 421;

Best Local Similarity 80.4%; Pred. No. 1.2e-149;

Matches 319; Conservative 18; Mismatches 43; Indels 17; Gaps 3;

QY 1 VKYGGSSVADAEIRRVAAEIRVATKQGNVVVVVSAMGDTTDLDLAQVCPAPPPR 60

Db 5 VQYGGSSVADADRIRRAERIVQTKQGNDIWVVVSAMGDTTDLDLAQVCPPEPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 120
Db 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 124
QY 121 LBEGRVVLVAGFGVSGDQTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVVGIFSDAP 180
Db 125 LDEGRVVLVAGFGVSGDQTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVVGIFSDAP 184
QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
Db 185 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 244
QY 241 IKDVPMEDEPILTGVAHDSRSEAKVTIVGLPDI PGYAAKVFRVAVARRRQRRHGAAB----- 295
Db 245 IKDVPMEDEPILTGVAHDSRSEAKVTIVGLPDI PGYAAKVFRVAVARRRQRRHGAAB----- 295
QY 296 RIQGRGRQDRHHLH---LPQTSRPPPKWTRSETRSASTOLLYDDHIGKVSIGAMR 352
Db 296 VLQNVSKVEDGKTDITFTCRSDSGPIAVAKLGLSLRDEIGFTQLLYDDHIGKVSIGAMR 355
QY 353 SHPGVTTATFCEALAAVGVNIELISTSEORSECCAT 389
Db 356 SHPGVTTATFCEALAAVGVNIELISTSEIRISVLCRDT 392

RESULT 2

US-08-532-828B-3
; Sequence 3, Application US/08532828B
; Patent No. 5686671
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: OGAWA, Yuri
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: TANAKA, Akiko
; APPLICANT: MATSUI, Hiroshi
; TITLE OF INVENTION: MUTANT ASPARTOKINASE GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P. C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS-DOS TEXT EDITOR
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,828B
FILING DATE: 27-OCT-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-101450
FILING DATE: 27-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-764-0 PCT
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC13869
US-08-532-828B-3
Query Match 55.8%; Score 1223.5; DB 1; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.1e-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADADRIRRAERIVQTKQGNDIWVVVSAMGDTTDLDLAQVCPPEPPAR 60
Db 5 VQYGGSSVADADRIRRAERIVQTKQGNDIWVVVSAMGDTTDLDLAQVCPPEPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 120
Db 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 124
QY 121 LBEGRVVLVAGFGVSGDQTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVVGIFSDAP 180
Db 125 LDEGRVVLVAGFGVSGDQTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVVGIFSDAP 184
QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
Db 185 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 244
QY 241 IKDVPMEDEPILTGVAHDSRSEAKVTIVGLPDI PGYAAKVFRVAVARRRQRRHGAAB----- 290
Db 245 IKDVPMEDEPILTGVAHDSRSEAKVTIVGLPDI PGYAAKVFRVAVARRRQRRHGAAB----- 290
QY 291 HGAERLQGRGRD-RHHLHLLP--QTSRPPPKWTRSETRSASTOLLYDDHIGKVSIL 347
Db 305 DGTDTITFTCPADGRAMEILKLVQVQ-----NW-----TNVLYDDQVGVKVSILV 350
QY 348 GAGMRSHPGVTTATFCEALAAVGVNIELISTSE 379
Db 351 GAGMRSHPGVTTATFCEALAAVGVNIELISTSE 382

RESULT 3

US-08-700-359-9
; Sequence 9, Application US/08700359
; Patent No. 5766925
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, MASAKAZU
; APPLICANT: USUDA, YOSHIHIRO
; APPLICANT: SUZUKI, TOMOKO
; APPLICANT: TANAKA, AKIKO
; APPLICANT: MATSUI, HIROSHI
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P. C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS-DOS TEXT EDITOR
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/700,359
FILING DATE: 08-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP-35019
FILING DATE: 04-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618

```
; REFERENCE/DOCKET NUMBER: 10-819-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-700-359-9

Query Match 55.8%; Score 1223.5; DB 1; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.le-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDAQQVCPPPPR 60
Db 5 VKYGGSSLESARIRVAERIVATKKAGNDVVVVCAMGDTTDELLELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGHNAKIIDVTPGRLQTA 120
Db 65 EMDMLLTAGERISNALVAMAIESLGAQAQSFSGAGVLTTERHGNARIVDVTPGRVREA 124
QY 121 LEEGRVVLVAGFQVGSQDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFGSADP 180
Db 125 LDEGKICITVAGFQVGNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYVADP 184
QY 181 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRGTVVVG 240
Db 185 RIVPNAQKLEKLSFEEMLELAAGSKILVRSYARAFNVPRVRSYSDNPGTLLIAGS 244
QY 241 IKDVPMDPILTVGAHDSRSEAKVTIVGLPIPGVAAKVFRAVA-----RRRQHR 290
Db 245 MEDIPVEEAVLTGVATDKSEAKVTLGSDKPGAQKVFRLADAEINIDMWLVQNVSSVE 304
QY 291 HGAERLQGRQD-RHHLHLLP--QTSGPPPKWNTSRSETRSASTQLLYDDHIGKVSLL 347
Db 305 DGTDTITFTCPRADGRAMEILKQLVQG-----TNVLYDDQVGVKVSIV 350
QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
Db 351 GAGMRSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 4
US-08-596-366-6
; Sequence 6, Application US/08596366
; Patent No. 5976983
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHOENOLPYRUVATE CARBOXYLASE.
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596.366
; FILING DATE: 29-APR-1996
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209775
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209776
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-153876
; FILING DATE: 05-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-784-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-596-366-6

Query Match 55.8%; Score 1223.5; DB 2; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.le-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDAQQVCPPPPR 60
Db 5 VKYGGSSLESARIRVAERIVATKKAGNDVVVVCAMGDTTDELLELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGHNAKIIDVTPGRLQTA 120
Db 65 EMDMLLTAGERISNALVAMAIESLGAQAQSFSGAGVLTTERHGNARIVDVTPGRVREA 124
QY 121 LEEGRVVLVAGFQVGSQDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFGSADP 180
Db 125 LDEGKICITVAGFQVGNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYVADP 184
QY 181 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRGTVVVG 240
Db 185 RIVPNAQKLEKLSFEEMLELAAGSKILVRSYARAFNVPRVRSYSDNPGTLLIAGS 244
QY 241 IKDVPMDPILTVGAHDSRSEAKVTIVGLPIPGVAAKVFRAVA-----RRRQHR 290
Db 245 MEDIPVEEAVLTGVATDKSEAKVTLGSDKPGAQKVFRLADAEINIDMWLVQNVSSVE 304
QY 291 HGAERLQGRQD-RHHLHLLP--QTSGPPPKWNTSRSETRSASTQLLYDDHIGKVSLL 347
Db 305 DGTDTITFTCPRADGRAMEILKQLVQG-----TNVLYDDQVGVKVSIV 350
QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
Db 351 GAGMRSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 5
US-08-967-104-6
; Sequence 6, Application US/08967104
; Patent No. 5919694
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHOENOLPYRUVATE CARBOXYLASE.
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
```

ADDRESSEE: P.C.
STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,104
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/596,366
FILING DATE: 29-APR-1996
APPLICATION NUMBER: JP 5-203775
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-203776
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-153876
FILING DATE: 05-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-784-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-104-6

Query Match 55.8%; Score 1223.5; DB 2; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.1e-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVWVVSAMGDTTDDLDAQQVCPAPPR 60
DB 5 VQYGGSSLESARIRNVAERIVATKKAGNDVVVVCAMGDTTDELELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSAQVITTTGHNAKIIDVTPGRLOTA 120
DB 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSAQVITTTGHNARIVDVTPGRVREA 124
QY 121 LSEGRVVLVAGFGVGSQDVKVTTTGRGSGDTTAVAMAAALGADVCEIYTDVDFGSADP 180
DB 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSGDTTAVAAAAALNADYCEIYSDVDGVYTDAP 184
QY 181 RIVRNARKLDVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
DB 185 RIVPNQAQLEKLSFEEMLELAAGVSKILVRSVEYARAFNPLVRSSYSNDPGLIAGS 244
QY 241 IKDVPMDPLTGVADHRSSEAKVTIVGLPDIIFYAAKVFRAVA-----RRRQHR 290
DB 245 MEDIPVEEAVLTGVATDKSEAKVTIVLIGSDKPGEAAKVFRALADAEINIDVQLNVSSVE 304
QY 291 HGAERLQGRGROD-RHLLHLFP--QTSGPPPKWNTSRSETSASTOLLYDDHIGKVSIL 347
DB 305 DGTDTITFTCFRAGRRAEILKKLQVQG-----NW-----TNVLYDDQGVKVSILV 350
QY 348 GAGMRSHPGVYATFCEALAAVGVNIELISTSE 379
DB 351 GAGMKSHPGVYATFCEALRDVNVNIELISTSE 382

RESULT 6
US-08-985-908-5
Sequence 5, Application US/08985908
Patent No. 6004773
GENERAL INFORMATION:
APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI NA
TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,908
FILING DATE: 05-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-325659
FILING DATE: 05-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-985-908-5

Query Match 55.8%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.1e-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVWVVSAMGDTTDDLDAQQVCPAPPR 60
DB 5 VQYGGSSLESARIRNVAERIVATKKAGNDVVVVCAMGDTTDELELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSAQVITTTGHNAKIIDVTPGRLOTA 120
DB 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSAQVITTTGHNARIVDVTPGRVREA 124
QY 121 LSEGRVVLVAGFGVGSQDVKVTTTGRGSGDTTAVAMAAALGADVCEIYTDVDFGSADP 180
DB 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSGDTTAVAAAAALNADYCEIYSDVDGVYTDAP 184
QY 181 RIVRNARKLDVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
DB 185 RIVPNQAQLEKLSFEEMLELAAGVSKILVRSVEYARAFNPLVRSSYSNDPGLIAGS 244
QY 241 IKDVPMDPLTGVADHRSSEAKVTIVGLPDIIFYAAKVFRAVA-----RRRQHR 290
DB 245 MEDIPVEEAVLTGVATDKSEAKVTIVLIGSDKPGEAAKVFRALADAEINIDVQLNVSSVE 304
QY 291 HGAERLQGRGROD-RHLLHLFP--QTSGPPPKWNTSRSETSASTOLLYDDHIGKVSIL 347
DB 305 DGTDTITFTCFRAGRRAEILKKLQVQG-----NW-----TNVLYDDQGVKVSILV 350
QY 348 GAGMRSHPGVYATFCEALAAVGVNIELISTSE 379
DB 351 GAGMKSHPGVYATFCEALRDVNVNIELISTSE 382

Db 351 GAGMKSHPGVTAFMEALRDVNVNIELISTSE 382

RESULT 8

US-08-985-916-5

Sequence 5, Application US/08985916

Patent No. 6221636

GENERAL INFORMATION:

APPLICANT: ATSUSHI HAYAKAWA, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI

APPLICANT: MASAKO IZUI, ATSUSHI HAYAKAWA, YASUHIKO YOSHIHARA, AND TSUYOSHI

APPLICANT: NAKAMATSU

TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR

CITY: ARLINGTON

STATE: VA

ZIP: 22152

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/985,916

FILING DATE: 05-DEC-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-325658

FILING DATE: 05-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 421 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-985-916-5

Query Match 55.8%; Score 1223.5; DB 3; Length 421;

Best Local Similarity 64.0%; Pred. No. 1.1e-116;

Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKQGNVTVVVSAMGDTTDDLLDAQVCPAPPP 60

Db 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELELAANVPVPP 64

Qy 61 ELDMLLTAGERISNALVAMAIESLGAAHARSFTGSQAGVITTTGTHGNAKIIVDTPGRLOTA 120

Db 65 EMDMLLTAGERISNALVAMAIESLGAAQSFTGSQAGVLTTERHGNARIVDTPGRVREA 124

Qy 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGADP 180

Db 125 LDEGKICIVAGFGVQNKETRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184

Qy 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVYARHNPVHVRSSYSRDPGTVVVGS 240

Db 185 RIVPNAQKLEKLSFEEMLELAAGSKILVRSVEYARAFNPLVRSSYNDPGLIAGS 244

Qy 241 IKDVPMDPILTVGAHDSSEAKVTIVGLPDIPGYAAKVFRVA-----RRRQHR 290

Db 245 MEDIPVEEAVLTGVATDKSEAKVTIVGLISDKPGEAAKVFRLADAENIDMVLQNVSSVE 304

Qy 291 HGAERLQGRQD--RHHLLIP--QTSPPPPKQWTRSETRASTOLLYDDHIGKVSLLI 347

Db 305 DGTDTITTCPRADGRRAVEILKKLQVQ-----TNVLYDDQVGKVSLLV 350

Qy 348 GAGMRSHPGVTATPCEALAAVGVNIELISTSE 379

US-08-852-730-14

Sequence 14, Application US/08852730

Patent No. 6090597

GENERAL INFORMATION:

APPLICANT: SEIKO HIRANO, MASAKAZU SUGIMOTO, EIICHI NAKANO,

APPLICANT: MASAKO IZUI, ATSUSHI HAYAKAWA, YASUHIKO YOSHIHARA, AND TSUYOSHI

APPLICANT: NAKAMATSU

TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER AND NEUSTADT, P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY

CITY: ARLINGTON

STATE: VA

ZIP: 22026

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/852,730

FILING DATE: 05-07-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-142812

FILING DATE: 05-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 421 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-852-730-14

Query Match 55.8%; Score 1223.5; DB 3; Length 421;

Best Local Similarity 64.0%; Pred. No. 1.1e-116;

Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKQGNVTVVVSAMGDTTDDLLDAQVCPAPPP 60

Db 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELELAANVPVPP 64

Qy 61 ELDMLLTAGERISNALVAMAIESLGAAHARSFTGSQAGVITTTGTHGNAKIIVDTPGRLOTA 120

Db 65 EMDMLLTAGERISNALVAMAIESLGAAQSFTGSQAGVLTTERHGNARIVDTPGRVREA 124

Qy 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGADP 180

Db 125 LDEGKICIVAGFGVQNKETRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184

Qy 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVYARHNPVHVRSSYSRDPGTVVVGS 240

Db 185 RIVPNAQKLEKLSFEEMLELAAGSKILVRSVEYARAFNPLVRSSYNDPGLIAGS 244

Qy 241 IKDVPMDPILTVGAHDSSEAKVTIVGLPDIPGYAAKVFRVA-----RRRQHR 290

Db 245 MEDIPVEEAVLTGVATDKSEAKVTIVGLISDKPGEAAKVFRLADAENIDMVLQNVSSVE 304

Qy 291 HGAERLQGRQD--RHHLLIP--QTSPPPPKQWTRSETRASTOLLYDDHIGKVSLLI 347

Db 305 DGTDTITTCPRADGRRAVEILKKLQVQ-----TNVLYDDQVGKVSLLV 350

Qy 348 GAGMRSHPGVTATPCEALAAVGVNIELISTSE 379

Db 351 GAGKSHPGVTAFFMEALRDVAVNIELISTSE 382
 RESULT 9
 US-08-532-828B-4
 ; Sequence 4, Application US/08532828B
 ; Patent No. 5689671
 ; GENERAL INFORMATION:
 ; APPLICANT: SUGIMOTO, Masakazu
 ; APPLICANT: OGAWA, Yuri
 ; APPLICANT: SUZUKI, Tomoko
 ; APPLICANT: TANAKA, Akiko
 ; APPLICANT: MATSUI, Hiroshi
 ; TITLE OF INVENTION: MUTANT ASPARTOKINASE GENE
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
 ; CITY: ARLINGTON
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: MS-DOS TEXT EDITOR
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/532,828B
 ; FILING DATE: 27-OCT-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 5-101450
 ; FILING DATE: 27-APR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: NORMAN F. OBLON
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 10-764-0 PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-413-3000
 ; TELEFAX: 703-413-2220
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 421 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Corynebacterium glutamicum
 ; STRAIN: AJ3463
 ; US-08-532-828B-4
 Query Match 55.6%; Score 1219.5; DB 1; Length 421;
 Best Local Similarity 63.8%; Pred. No. 2.8e-116;
 Matches 250; Conservative 55; Mismatches 60; Indels 27; Gaps 5;
 QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQQVCAPP 60
 Db 5 VQYGGSSLESABIRIRVAERIVATKKAGNDVVVVSAMGDTTDLLELAAAVNP 64
 QY 61 ELDMLTAGERISNALVAMAIESLGHAHSFTGSOAGVITTHGNAKIIDVTPGRLOTA 120
 Db 65 EMDMLTAGERISNALVAMAIESLGHAQSFTGSOAGVITTHGNARIVDTPGRVREA 124
 QY 121 LEEGRVVLVAGFGVGSQDTKDVTTLGRGSDTTAVAAAAALGADVCEIYTDVDFGSADP 180
 Db 125 LDEGKICIVAGFGVNKETRDVTTLGRGSDTTAVAAAAALNADVCEIYSDVDGVYADP 184
 QY 181 RIVNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSSYSYDRPGTVVVG 240

Db 185 RIVNACKLEKLSFEEMLELAAGVKSKILVLRSEYARAFNPLVRSSYSNDPGTTLIAGS 244
 QY 241 IKQVPMEDPILTVANDRSEAKVTIVGLDIPGAAKVRPAVA-----RRRQHR 290
 Db 245 MEDIPVEEAVLTGATDKSEAKVTIVGISDKPGETAKVFRALADAENIDMVLQNVSSVE 304
 QY 291 HGAERLQGRGROD-RHHLHLP--QTSGPPPKWNRSETSRSASTQLLYDDHIGKVSII 347
 Db 305 DGIITDITFTCPRADGRAMEILKLVQV-----NW-----INVLYDDQGVKSLV 350
 QY 348 GAGMRSHPGVTATFCALAAVGVNIELISTSE 379
 Db 351 GAGKSHPGVTAFFMEALRDVAVNIELISTSE 382
 RESULT 10
 US-09-252-991A-29720
 ; Sequence 29720, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 29720
 ; LENGTH: 461
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-29720
 Query Match 37.5%; Score 823; DB 4; Length 461;
 Best Local Similarity 42.6%; Pred. No. 1.3e-75;
 Matches 170; Conservative 80; Mismatches 105; Indels 44; Gaps 4;
 QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQQVCAPP 60
 Db 54 VQKEGTSVGTVERIEQVAEKVKFREAGDDVVVVVSAMSGETNRLIGLANQIMEQV 113
 QY 61 ELDMLTAGERISNALVAMAIESLGHAHSFTGSOAGVITTHGNAKIIDVTPGRLOTA 120
 Db 114 ELDMVSTGEQVTIALLSMALIKRGVPAVSYTGQVRILTDSHTKARILHDDTHRAD 173
 QY 121 LEEGRVVLVAGFGVGSQDTKDVTTLGRGSDTTAVAAAAALGADVCEIYTDVDFGSADP 180
 Db 174 LKAGRVVVVAGFQGV-DGNGNITTLGSGSDTTGVALAALKADECIYTDVDSVTTDP 232
 QY 181 RIVNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSSYSYDRPGTVVVG 240
 Db 233 RVVQPARLDKITFEEMLEMASLGSKVLQIRAVEFAGKYNVPLRVLHLSFOEGFTLITD 292
 QY 241 IKQVPMEDPILTVANDRSEAKVTIVGLDIPGAAKVRPAVA-----RRRQ 288
 Db 293 DEESMEQPIISGIAFNDRDEAKUTIRGVPDTPGVAFKILGPISAANVEVDIMVQVADHN 352
 QY 289 -----HRHGAERLQGRGRODRHHLHLLPOTSGPPPKWNRSETSRSASTQLLYD 340
 Db 353 TDTFTFVHRNDVNLAE-----ILKQTA-----ANTIGAREALGDTN 389
 QY 341 IGKVSLLGAGMRSHPGVTATFCALAAVGVNIELISTSE 379
 Db 390 IAKVSIQVGMRSAGVAVSRMFEALAKESINIQMISTSE 428
 RESULT 11
 US-07-684-135A-2
 ; Sequence 2, Application US/07684135A

Patent No. 5243039
GENERAL INFORMATION:
APPLICANT: Schendel, Frederick J. and
APPLICANT: Flickinger, M. C.
TITLE OF INVENTION: Bacillus MGA 3 Aspartokinase
TITLE OF INVENTION: II Gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 5243039west Center
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb
COMPUTER: No. 5243039thgate 386
OPERATING SYSTEM: DOS 4.0
SOFTWARE: Wordperfect- 5.0
CURRENT APPLICATION DATA: US/07/684,135A
APPLICATION NUMBER: US/07/684,135A
FILING DATE: 19910412
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 600.216-US-01
TELEPHONE: (612) 332-5300
TELEFAX: (612) 332-9081
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 411 amino acids
TYPE: AMINO ACID
TOPOLOGY: Linear
MOLECULE TYPE: Polypeptide
FEATURE:
NAME/KEY: Aspartokinase II ~ dimer subunit
LOCATION: 1 to 411
US-07-684-135A-2

Query Match 35.8%; Score 785; DB 1; Length 411;
Best Local Similarity 42.1%; Pred. No. 8.4e-72;
Matches 170; Conservative 81; Mismatches 97; Indels 56; Gaps 6;
QY 1 VKYGGSSVADAERIRRAERIVATKKGNDVVVWSAMGDTTDDLLDQAQVCPAPPPR 60
DB 5 VKFGGTSVGSVERILNVANRIVIEKKNQNDVVVWSAMGKTTDELVDLAKQISAHPPKR 64
QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIDVTGRLQTA 120
DB 65 EMDMLTTGGEQVITSLALWALNEKGYEAI SYTGWQAGITTEPVFGNARILNIETEKIQK 124
QY 121 LEEGRVVLVAGFQGVSDTKVTTLGRGSDTTAVAMAALGADVCEIYTDVDGIFSDAP 180
DB 125 LNEGRIVVVGFGQGDIEH-GEITTLGRGSDTTAVALAALKAERKCDIYTDVGVFTTDP 183
QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVVRSSYSYSDRPGTVVGS 240
DB 184 RYVKSARKLASISYDEMELANGLGVLHRAVEFAKNGITLVRSSMEREGTIE-- 241
QY 241 IKDVEPE-DPILTVGVAHDRSEAKVTIPLGPYAAKVFRAVARRRRQRHGAERLQG 299
DB 242 -EETVMEQNLVVRGVARDEITRVTVFGPLNSLTSLSTIFTLA----- 284
QY 300 RCRDRPHLHLLPQSGPPPKWNRSETRASATQLLYDD----- 339
DB 285 ---QNRINVDIIQSA-----TDAETNLSFSIKSDSDDLEETWAVLENNKLLNYQGI 333
QY 340 ---HIGKVLISAGMRGHPGVTAFCALAAVGNIELISTSE 379
DB 334 ESETGLAKVISVGSWMISNPGVAQWFEVLALNGIQVQWVSTSE 377

RESULT 12

US-09-056-556-227
Sequence 227, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 227:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-056-556-227

Query Match 35.1%; Score 770; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.5e-71;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 QKYGGSSVADAERIRRAERIVATKKGNDVVVWSAMGDTTDDLLDQAQVCPAPPPRE 61
DB 1 QKYGGSSVADAERIRRAERIVATKKGNDVVVWSAMGDTTDDLLDQAQVCPAPPPRE 60
QY 62 LDMLLTAGRISNALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIDVTGRLQTA 121
DB 61 LDMLLTAGRISNALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIDVTGRLQTA 120
QY 122 EGRVVLVAGFQGVSDTKVDVITLGRGSDTTAVAM 157
DB 121 EGRVVLVAGFQGVSDTKVDVITLGRGSDTTAVAM 156

RESULT 13

US-09-072-596-222
Sequence 222, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.

APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 222:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-072-596-222

Query Match 35.1%; Score 770; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.5e-71;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 QKYGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPRE 61
DB 1 QKYGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPRE 60
QY 62 LDMLLTAGERISNALVAMAEISGAHARSFTGSQAGVITGTHGNAKIIDVTPGRLQAL 121
DB 61 LDMLLTAGERISNALVAMAEISGAHARSFTGSQAGVITGTHGNAKIIDVTPGRLQAL 120
QY 122 EGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAM 157
DB 121 EGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAM 156

RESULT 14
US-09-328-352-6511
Sequence 6511, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6511
LENGTH: 438
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6511

Query Match 34.0%; Score 746.5; DB 4; Length 438;
Best Local Similarity 42.0%; Pred. No. 8.2e-68;
Matches 168; Conservative 73; Mismatches 124; Indels 35; Gaps 8;

QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB 17 VQYGGTSMGTPERILNVAERVKRWHDHGHKVVVVSAMSGETNRLALAKALETDPDR 76
QY 61 ELDMLLTAGERISNALVAMAEISGAHARSFTGSQAGVITGTHGNAKIIDVTPGRLQTA 120
DB 77 ELDQMVSTGEQVTISMLAVALSIGVEAKSYTGROVGKIDTSFTKARIESIDTVMTND 136
QY 121 LEBGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVIGISADP 180
DB 137 LDAGRIVVAGFQGVSDAN-GNTTTLGRGSGDTSGVAIAAALKADECIYTDVGVVYTD 195
QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMURCQVEYARRHNPVHVRSS----- 228
DB 196 RVAPKAKKIDRISFEEMLEMAAGSKVLQIRSVFPAKQVQVPLVSSPDNDGDAFDD 255
QY 229 YSDRPGTVVVGSIKDVPMEDPILTVVAHDSKAKVTIVGLPDIPGYAAKVFRAVARR--- 285
DB 256 FKENVGTLLITAEADT-MEQPIIAGIAFNRDEAKLITLGVDPFGIASKILSPVSDANIE 314
QY 286 -----RQHRHGAERLQGRGDRHHLH-LLPQTSQPPPKWTRSETESASTQLLYDD 339
DB 315 VDMIVQNVBEDGTTDTFTTVNRVDLAKAEKILNETA-----KN---IGAREVSTR---D 362
QY 340 HIGKVSILGAGMRSHPGVTTATFCEALAAVGVNIELISTSE 379
DB 363 DIVKVSIVGVGMRSHAGVASKMFTALADEGINILMISTSE 402

RESULT 15

US-09-134-001C-5225
Sequence 5225, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5225
LENGTH: 424
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5225

Query Match 25.7%; Score 563; DB 4; Length 424;
Best Local Similarity 35.6%; Pred. No. 4.8e-49;
Matches 138; Conservative 82; Mismatches 142; Indels 26; Gaps 8;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB 29 VLKPGSSVSDFKKIRKINIAEMKLTREDGGEELIVVVSAMGKTTDQLMNVSSLTSTPKDQ 88
QY 61 ELDMLLTAGERISNALVAMAEISGAHARSFTGSQAGVITGTHGNAKIIDVTPGRLQTA 120
DB 89 ELALLTTGEGQVTSVLSVNLNDIGVNAKAMTGYAGIKTVGHHLKSKIAEINPINEA 148

QY 121 LEBGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVIGISADP 180
DB 149 FKNDILVWAGFQGVSDNF-ELTTLGRGSGDTTAVALAAS-NQTPCEIYTDVGVVATDP 206
QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMURCQVEYARRHNPVHVRSSYSRDRGTVVVGS 240
DB 207 RIHNEAKLEYSVEENWMSALGAGVLETRSELAKNYDIPLYLGTLSNVKGTWIMS- 265
QY 241 IKDVPMDPILTVVAHDSKAKVTI-VGLDIPGYAAKVFRAVARRRRRHRHGAERLQ 299

```
8
Db 266 -KSDLEKKAVTGVALDTHMHVITISYPLPD-NQLLTQLFTAL-EEESVNVDMISQIVNL 322
Qy 300 RGRQ-----DRHHLHLLPCTSGPPPWKNWTRSETRSASTOLLYDDDHGKVSLLICAGM 351
Db 323 EGLQLSFSIKSDAHQISSILE-----NLSTHFSALDYKINEAYVKISLIGSGM 371
Qy 352 RSHPGVTATFCEALAAVGVNIELISTSE 379
Db 372 RMSGVASKAFTTLINSDIPFYQTTTSE 399
```

Search completed: November 21, 2003, 16:11:37
Job time : 9.9415 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 16.537 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-4

Perfect score: 2193

Sequence: 1 VQYGGSSVADAEIRRAEIRVAE.....SAATRRPCTGRGDRWACQ 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 656188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1615.5	73.7	421	9	US-09-793-306-163
2	1224.5	55.8	421	10	US-09-738-626-3781
3	1224.5	55.8	421	11	US-09-746-660A-24
4	1223.5	55.8	421	15	US-10-226-136-5
5	1217.5	55.5	421	15	US-10-067-974-2
6	1211.5	55.2	421	15	US-10-067-974-16
7	1144.5	52.2	430	15	US-10-156-761-12093
8	770	35.1	156	12	US-10-084-843-227
9	770	35.1	156	12	US-10-193-002-222
10	673	30.7	394	12	US-09-882-227-290
11	377	17.2	564	10	US-09-890-813-17
12	356	16.2	555	10	US-09-890-813-8
13	350	16.0	560	10	US-09-890-813-6
14	314.5	14.3	262	12	US-10-301-997-23
15	310.5	14.2	449	11	US-09-989-339-23

ALIGNMENTS

RESULT 1

US-09-793-306-163
; Sequence 163, Application US/09793306
; Patent No. US2000098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Ovendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 163
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: MO-2 (aspartokinase)
US-09-793-306-163

Query Match 73.7%; Score 1615.5; DB 9; Length 421;
Best Local Similarity 85.1%; Pred. No. 1.9e+146;
Matches 338; Conservative 7; Mismatches 35; Indels 17; Gaps 3;
Qy 1 VQYGGSSVADAEIRRAEIRVAEIRVATKKQGNVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
Db 5 VQYGGSSVADAEIRRAEIRVATKKQGNVVVVVSAMGDTTDDLLDLAQVCPAPPPR 64
Qy 61 ELDMLTAGERISNALVAMAEISLGAHARSFTGSQAGVITGTHGNAKIIDVTFGLQTA 120

Sequence 16, Appl
Sequence 7, Appl
Sequence 34, Appl
Sequence 12, Appl
Sequence 14293, A
Sequence 518, Ap
Sequence 6362, Ap
Sequence 34, Appl
Sequence 34, Appl
Sequence 40, Appl
Sequence 88, Appl
Sequence 88, Appl
Sequence 60, Appl
Sequence 60, Appl
Sequence 239, Appl
Sequence 10162, A
Sequence 10, Appl
Sequence 114, App
Sequence 114, App
Sequence 160, App
Sequence 40, Appl
Sequence 160, App
Sequence 3, Appl
Sequence 276, App
Sequence 6, Appl
Sequence 169, App
Sequence 4, Appl
Sequence 17, Appl

Db 65 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGHNAKIIDVTPGSLQTA 124
 QY 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVVGIFSDAP 180
 Db 125 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVVGIFSDAP 184
 QY 181 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 240
 Db 185 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 244
 QY 241 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDPICGYAAKVFRAVARRRRQHGAAE----- 295
 Db 245 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDPICGYAAKVFRAVARRRRQHGAAE----- 295
 QY 296 RLQGRGRQRHLLHLLPQTS---GPPPKWNWTRSETRASTOLLYDDHIGKYSLLIGAGMR 352
 Db 296 VLQNVSKVEDGKTIDTFTCSRDPGPAAVKLSLRNEIGFSLYDLDHIGKYSLLIGAGMR 355
 QY 353 SHPGVTATFCEALAAVGVNIELISTSEDQRSCCAAT 389
 Db 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCDRT 392

RESULT 2

US-09-738-626-3781
 ; Sequence 3781, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOUCUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAKO
 ; APPLICANT: SENO, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OKAZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; PRIOR FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO 3781
 ; LENGTH: 421
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-3781

Query Match 55.8%; Score 1224.5; DB 10; Length 421;
 Best Local Similarity 64.0%; Pred. No. 6.9e-109;
 Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
 QY 1 VKYGGSSVADAEIRIRVAERIVATKQGNDDVVVVSAMGTTTDDLLDLAQVCPAPP 60
 Db 5 VKYGGSSLESERIRNVAERIVATKAGNDVVVVCSSAMGTTTDELLEAAVNPVPPAR 64
 QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGHNAKIIDVTPGSLQTA 120
 Db 65 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGHNAKIIDVTPGSLQTA 124
 QY 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVVGIFSDAP 180
 Db 125 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVVGIFSDAP 184

QY 181 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 240
 Db 185 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 244
 QY 241 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDPICGYAAKVFRAVARRRRQHGAAE----- 295
 Db 245 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDPICGYAAKVFRAVARRRRQHGAAE----- 295
 QY 296 RLQGRGRQRHLLHLLPQTS---GPPPKWNWTRSETRASTOLLYDDHIGKYSLLIGAGMR 352
 Db 296 VLQNVSKVEDGKTIDTFTCSRDPGPAAVKLSLRNEIGFSLYDLDHIGKYSLLIGAGMR 355
 QY 353 SHPGVTATFCEALAAVGVNIELISTSEDQRSCCAAT 389
 Db 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCDRT 392

RESULT 3

US-09-746-660A-24
 ; Sequence 24, Application US/09746660A
 ; Publication No. US20030049804A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pompejus, Markus
 ; APPLICANT: Kroger, Burkhard
 ; APPLICANT: Schroder, Hartwig
 ; APPLICANT: Zelder, Oskar
 ; APPLICANT: Haberhauer, Gregor
 ; APPLICANT: Kim, Jun-Won
 ; APPLICANT: Lee, Heung-Schick
 ; APPLICANT: Hwang, Byung-Joon
 ; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
 ; FILE REFERENCE: BGI-121CP2
 ; CURRENT APPLICATION NUMBER: US/09/746,660A
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 09/606740
 ; PRIOR FILING DATE: 2000-06-23
 ; PRIOR APPLICATION NUMBER: 09/603124
 ; PRIOR FILING DATE: 2000-06-23
 ; PRIOR APPLICATION NUMBER: 60/141031
 ; PRIOR FILING DATE: 1999-06-25
 ; PRIOR APPLICATION NUMBER: 60/142101
 ; PRIOR FILING DATE: 1999-07-02
 ; PRIOR APPLICATION NUMBER: 60/148613
 ; PRIOR FILING DATE: 1999-08-12
 ; PRIOR APPLICATION NUMBER: 60/187970
 ; PRIOR FILING DATE: 2000-03-09
 ; PRIOR APPLICATION NUMBER: DE 19931420.9
 ; PRIOR FILING DATE: 1999-07-08
 ; NUMBER OF SEQ ID NOS: 125
 ; SOFTWARE: PatentIn Vers. 2.0
 ; SEQ ID NO 24
 ; LENGTH: 421
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-746-660A-24

Query Match 55.8%; Score 1224.5; DB 11; Length 421;
 Best Local Similarity 64.0%; Pred. No. 6.9e-109;
 Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
 QY 1 VKYGGSSVADAEIRIRVAERIVATKQGNDDVVVVSAMGTTTDDLLDLAQVCPAPP 60
 Db 5 VKYGGSSLESERIRNVAERIVATKAGNDVVVVCSSAMGTTTDELLEAAVNPVPPAR 64
 QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGHNAKIIDVTPGSLQTA 120
 Db 65 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGHNAKIIDVTPGSLQTA 124
 QY 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVVGIFSDAP 180
 Db 125 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVVGIFSDAP 184

QY 291 HGAERLQGRQD-RHHLLLP--QTSPPPWKNWTRSETRSASTQLLYDDHIGKVSLL 347
DB 305 DGTDTITFTCPDGRAMEILKKLVQVQ-----NW-----TNVLYDDQVKKVSLV 350
QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISISE 382

RESULT 6

US-10-067-974-16
; Sequence 16, Application US/10067974
; Publication No. US20030055232A1
; GENERAL INFORMATION:
; APPLICANT: Li, Liang-Yew
; APPLICANT: Irei, Kelli J.
; TITLE OF INVENTION: Polynucleotide Constructs for Increased Lysine Production
; FILE REFERENCE: 1533.2640001
; CURRENT APPLICATION NUMBER: US/10/067,974
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,183
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 16
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus Sequence of Protein Sequence Alignment
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (40)..(40)
; OTHER INFORMATION: May be either Cys or Val
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (317)..(317)
; OTHER INFORMATION: May be either Ser or Ala
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (345)..(345)
; OTHER INFORMATION: May be either Gly or Asp
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (380)..(380)
; OTHER INFORMATION: May be either Thr or Ile
US-10-067-974-16

Query Match 55.2%; Score 1211.5; DB 15; Length 421;
Best Local Similarity 63.5%; Pred. No. 1.2e-107; Mismatches 61; Indels 27; Gaps 5;
Matches 249; Conservative 55;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLLDLAQVCPAPPP 60
DB 5 VQYGGSSLESARIRVAERIVATKKAGNVVVSXSAMGDTTDELLELAAAVNPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITGTGHNAKIIDVTPGRLOTA 120
DB 65 EMDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITGTGHNARIVDVTGKRVREA 124
QY 121 LEEGRVVLVAGFGQVSQDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVGI FSA DP 180
DB 125 LDEGKICIVAGFGQVKNKEDVTTLGRGSDTTAVALLAALNADVCEIYSDVDGVY TAD P 184
QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSYSDRPGT VVGS 240
DB 185 RIVPNAOKLEKLSFEEMLEMAAGSKVLRLSRVEYAFNPLRVSRSSYNPDGTLIAGS 244
QY 241 IKDVPMDPILTVGAHDRSEAKVTIVGLPDIPGYAAKVFRAVRA--RRRQR 290
DB 245 MEDIPVEAVLTGVATDKSAKVTVLGISDKPGEAAKVFEALADAEINIDMVLQNVSSVE 304
QY 291 HGAERLQGRQD-RHHLLLP--QTSPPPWKNWTRSETRSASTQLLYDDHIGKVSLL 347

DB 305 DGTDTITFTCPDGRAMEILKKLVQVQ-----NW-----TNVLYDDQVKKVSLV 350
QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISIXSE 382

RESULT 7

US-10-156-761-12093
; Sequence 12093, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12093
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12093

Query Match 52.2%; Score 1144.5; DB 15; Length 430;
Best Local Similarity 60.2%; Pred. No. 3.4e-101; Mismatches 75; Indels 29; Gaps 4;
Matches 239; Conservative 54;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLLDLAQVCPAPPP 60
DB 5 VQYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDELIDLAEQVSPMESGR 64
QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITGTGHNAKIIDVTPGRLOTA 120
DB 65 EFDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITGDSVHNKARIIDVTPGRIRTA 124
QY 121 LEEGRVVLVAGFGQVSQDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVGI FSA DP 180
DB 125 LDEGNIAIVAGFGQVSQDKKDIITLGRGSDTTAVALLAALDAEVCEIYTDVGVFTAD P 184
QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSYSDRPGT VVGS 240
DB 185 RVKKAKKIDWIAFEDMLELAAGSKVLLHRCVEYARRYNIPHRSSFSGLQGTWV--- 241
QY 241 IKDVPMDPILTVGAHDRSEAKVTIVGLPDIPGYAAKVFRAVRA-- 286
DB 242 -SNTPLVQKAOQCEQVEQAIISGVAHDTSEAKVTIVGVDPKPGEAASIFRAIDA EAVN 300
QY 287 ----RQRHGAERLQGRQDRHLLHLLPQTSGPPPWKNWTRSETRSASTQLLYDDHIG 342
DB 301 IDMVQNVNSAATGLTDIS-----FTLPKTEGRKAIDALEKAKSVIGFDSLR YDDQIG 353
QY 343 KVSLLIGAMRSHPGVTATFCEALAAVGVNIELISTSE 379
DB 354 KISLVGAGMKNTPGVTAGFPEALSDAGVNIELISTSE 390

RESULT 8

US-10-084-843-227
; Sequence 227, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/05/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 227:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 227:

US-10-084-843-227
Query Match 35.1%; Score 770; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.4e-66;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QKYGSSVADAEIRIRVAERIVATKKQGNDDVVVWSANGDTTDDLLDLAQVCAPPPRE 61
Db 1 QKYGSSVADAEIRIRVAERIVATKKQGNDDVVVWSANGDTTDDLLDLAQVCAPPPRE 60

QY 62 LDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTAL 121
Db 61 LDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTAL 120

QY 122 EGRVVLVAGFGVQSQDTKDVTTTLGRGSDTTAVAM 157
Db 121 EGRVVLVAGFGVQSQDTKDVTTTLGRGSDTTAVAM 156

RESULT 9
US-10-193-002-222
Sequence 222, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS

NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 222:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 222:

US-10-193-002-222
Query Match 35.1%; Score 770; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.4e-66;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QKYGSSVADAEIRIRVAERIVATKKQGNDDVVVWSANGDTTDDLLDLAQVCAPPPRE 61
Db 1 QKYGSSVADAEIRIRVAERIVATKKQGNDDVVVWSANGDTTDDLLDLAQVCAPPPRE 60

QY 62 LDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTAL 121
Db 61 LDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTAL 120

QY 122 EGRVVLVAGFGVQSQDTKDVTTTLGRGSDTTAVAM 157
Db 121 EGRVVLVAGFGVQSQDTKDVTTTLGRGSDTTAVAM 156

RESULT 10
US-09-882-227-290
Sequence 290, Application US/09882227
Publication No. US20030158396A1
GENERAL INFORMATION:
APPLICANT: Kleanthous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Oocmen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides

```

; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 290
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-290

Query Match      30.7%; Score 673; DB 12; Length 394;
Best Local Similarity 40.1%; Pred. No. 5.8e-56;
Matches 151; Conservative 83; Mismatches 119; Indels 24; Gaps 6;

Qy 13 ERIRVAERIVATKQGNDDVVVVVVSAMGDTTDDLDAQQVCPAPPRELDMLLTAGERI 72
Db 5 ERHNVAQRVLESVTIQLHQVVVVVSMSGETDLLEFQKNFHNPNKREMDRIVSGELV 64

Qy 73 SNALVAMAIESLGAHARSFTGSAQVITTTGTHGNAKIIDVTPRLQALBEGRVVLVAGF 132
Db 65 SSAALSVALERYGHRAISLSGKEAGILTSHFQNAVIGSIDTKRITELLEKNYIVVIAGF 124

Qy 133 QGVSDQTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGIFSDADPRIVNARKLDTV 192
Db 125 QG-ADTQGETTTLGRGSDLSVALAGALKHCEIYTDVGVYITDPRIBKAKIAQI 183

Qy 193 TFEEMLEMAACGAKVLMRCVYARHNPVHVRSSYSDRPGTVVVG--SIRKDVPMEDPI 250
Db 184 SYDMLSLAMGAKVLLNSVELAKXLSVKLVTRNSFNHSEGLTVAEKDFKGERMETPI 243

Qy 251 LTGVANDRSEAKTVIQLDPIDPYAAKVFRAVARRRQRHGAERLQGRGRQDRHHLHL 310
Db 244 VSGIALDKQARVSMGVEDRPGIAAEIFCALA---EYRINVDIMVQITIGDGTDLDF 299

Qy 311 -LPQSGPPPKWTRSETRRSASTQL-----YDDHKGKVSLLGAGMRSHPGVTATFC 362
Db 300 TIVKTO-----TEEKQALKPFLQWDSIDYDENIAKVISGVGMKSHGVSATAP 350

Qy 363 EALAAVGVNIELSTSE 379
Db 351 KALAKDNINIMISTSE 367

RESULT 11
US-09-890-813-17
; Sequence 17, Application US/09890813
; Publication No. US20020183486A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; TITLE OF INVENTION: Aspartate Kinase
; FILE REFERENCE: BB1430 PCT
; CURRENT APPLICATION NUMBER: US/09/890,813
; CURRENT FILING DATE: 2001-08-02
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Glycine max
US-09-890-813-17

Query Match      17.2%; Score 377; DB 10; Length 564;
Best Local Similarity 27.0%; Pred. No. 2.6e-27;
Matches 128; Conservative 71; Mismatches 165; Indels 110; Gaps 13;

Qy 1 VQYGGSSVADAEIRRAERIVATKQGNDDVVVVVSAMGDTTDDL-----47

```

```

Db 87 VMKFGSSVASADRMKEVATLILSFPEE--RPVIVLSAMGKTTNKLLLAGKAVSCGVIN 144
Qy 48 -----DLAQQVCPAPPREL-----DMLLTAG 69
Db 145 VSSTEEELCFIKDLHLRTVDQLGVDSGVISKHLEQLLKGIAAMKELTKRTQDYLVSFG 204
Qy 70 ERISNALVAMAIESLGAHARSFTGSAQVITTTGTHGNAKIIDVT---FRLQ-TALEEG 124
Db 205 ECMSTRIFAAYLNKIGVKARQYDAFEIGFITTTDDFTNADILEATYPAVAKRLHGDWSDP 264
Qy 125 RVVLVAGFQGVSDQTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGIFSDADPRIVR 184
Db 265 AIAIVTGLGKARKSCAVTTILGRGSDLTATTIGKALGLPEIQVWKDVGDLTCDPNIYP 324
Qy 185 NARKLDTVTFEEMLEMAACGAKVLMRCVYARHNPVHVRSSYSDR-PGTVVVGSIKD 243
Db 325 KAEFVPYLTTFDEAAELAYFGAQLHPSQMRPARESDIPVRKNSYNPKAPGTLIT---KA 381
Qy 244 VPMEDPILTGVANDRSEAKTVIQLDPIDP--GYAAKVFRAVARRRQRHGAERLQGRG 301
Db 382 RDMSKAVLTISVLKRNVTMLDIASTRMLGOYGLAKVFSIFEELGISVDVWATSEVSVSL 441
Qy 302 QDRHHLHLLPQSGPPPKWTRSETRRSASTOLLVDDHIGK-----VS 345
Db 442 TLD-----PSKLWSRELITQASL-----DHVVEELEKIAVNNLQNESIIS 483
Qy 346 LIGAGMRSHPGVTATFCEALAAVGVNIELIS-----TSEBQRGCCAA 388
Db 484 LIGNVQRSSL-ILERLSRVLTGLVTQVMISQASKVNSLVNDSEAEQCPRA 536

RESULT 12
US-09-890-813-8
; Sequence 8, Application US/09890813
; Publication No. US20020183486A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; TITLE OF INVENTION: Aspartate Kinase
; FILE REFERENCE: BB1430 PCT
; CURRENT APPLICATION NUMBER: US/09/890,813
; CURRENT FILING DATE: 2001-08-02
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Zea mays
US-09-890-813-8

Query Match      16.2%; Score 356; DB 10; Length 555;
Best Local Similarity 29.9%; Pred. No. 2.7e-25;
Matches 101; Conservative 57; Mismatches 116; Indels 64; Gaps 7;

Qy 1 VQYGGSSVADAEIRRAERIVATKQGNDDVVVVVSAMGDTTDD-----44
Db 86 VMKFGSSVSSAARMAEVAGLITTFPEE--RPVVVLSAMGKTTNNLLLAGKAVGCGVIH 143
Qy 45 -----DLDLAQQVCPA-----PPPRELDMLLTAG 69
Db 144 VSETEENWNVKSLHIKTVDELGLPRSVIQDMLDELEQLLKGIAAMKELTPRTSDYLVSFG 203
Qy 70 ERISNALVAMAIESLGAHARSFTGSAQVITTTGTHGNAKIIDVTPRELQAL-----EEG 124
Db 204 ECMSTRIFAAYLNKIRVKARQYDAFDIGFITTTDFGNADILEATYPAVAKRLHGDWIDQP 263
Qy 125 RVVLVAGFQGVSDQTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGIFSDADPRIVR 184
Db 264 AIPVVTGLGKWKSGAVTTILGRGSDLTATTIGKALGLREIQVWKDVGDLTCDPNIYP 323
Qy 185 NARKLDTVTFEEMLEMAACGAKVLMRCVYARHNPVHVRSSYSDR-PGTVVVGSIKD 243

```

```

; Publication No. US20030148346A1
; GENERAL INFORMATION:
; APPLICANT: Holden, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Garstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/301,997
; FILING DATE: 22-No. US20030148346A1-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/527,431
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US/08/887,534
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Rip-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
; US-10-301-997-23

Query Match 14.3%; Score 314.5; DB 12; Length 262;
Best Local Similarity 33.9%; Pred. No. 8.7e-22;
Matches 87; Conservative 47; Mismatches 82; Indels 41; Gaps 8;

Qy 140 KDVTTILGRGSDTTAVAMAALGADVCEIYTDVDFGFSADPRIVRNARKLDTVTREEMLE 199
Db 5 QEUTTLGRGSDTTAVALAVS-NQIPCEIYTDVGVATDPLLPKAKRELDIVSYEEMW 63
Qy 200 MAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVVGSIKDVPEDPILTVGAHRS 259
Db 64 MSALGAGVLETSVELAKVNPPLYLGKTLNVKGTWMSN--EELEKKAVTGVALDKH 121
Qy 260 EAKVTI-VGLPDIPGVAAKVFAVARRRQHRHGAEE-----FLQR-----GROD 304
Db 122 MMHVTISYFLPD-----NQILTQLFTEBEGAVNVDMISQIVNLDDGLQLSFTTKDSD 173
Qy 305 RHHLHLLPQTSGPPPKWNTSRSTRSASTQLLY--DDHIGKVSILGAGMRSHPGVTATFC 362
Db 174 FHOISMLT-----LNQYEALEYKINEHYVKISLIGSGNRDMSGVASKAF 220
Qy 363 EALAAVGVNIELISTSE 379
Db 221 LTLIENNIPFYQTTTSE 237

RESULT 15
US-09-989-339-23
; Sequence 23, Application US/09899339
; Publication No. US2003008886A1
; GENERAL INFORMATION:

```

```

; Publication No. US20030148346A1
; GENERAL INFORMATION:
; APPLICANT: Holden, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Garstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/301,997
; FILING DATE: 22-No. US20030148346A1-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/527,431
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US/08/887,534
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Rip-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
; US-10-301-997-23

Query Match 16.0%; Score 350; DB 10; Length 560;
Best Local Similarity 26.4%; Pred. No. 1e-24;
Matches 120; Conservative 76; Mismatches 155; Indels 104; Gaps 12;

Qy 1 VOKYGGSSVADEAIRRVAERIVATKQGNVNVVVSAMGDTTDDLLDLAQ----- 52
Db 86 VMKFGSSVSAARMAEAVGLILTPPE--RPVVVLSAMGKTTNNLLAGEKAVGCGVIH 143
Qy 53 -----VCPAP-----PPRELDMLLTAGE 70
Db 144 VSEIEEWMNVKSLHKITVDELGLXICNTSYELEQLLKGIAAMKELTPRTSDVLSFGE 203
Qy 71 RISNALVAMATIESLGAHARSTGSGAGVITGTHGNAKIIDVTPCRLQAL-----BEGR 125
Db 204 CMSTRIFSAYLKIRVKARQYDAFDIGFITTFEGNADILEATYPAVAKRLHGDWIDQPA 263
Qy 126 VVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGFSADPRIVRN 185
Db 264 IPVVTGFLGKWKSGAVTTLGRGSDLTATTIGKALGLREIQWKVDVGLTCDPNLYPH 323
Qy 186 ARKLDVTTFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDR--PGTVVVVGSIKDV 244
Db 324 AKTVPYLTFEEATELAYFGAQLHPOSMPAREGDIPIRVKNSYNPKAPGTLITRQ-RDM 382
Qy 245 PMEDPILTVGAHRSSEAKVTIVGLPDIPGVAAKVFAVARRRQHRHGAERLOGRGOD 304
Db 383 DXGLVVLTSIV---LKSNTVMTLD-----IVSTRMLGQYGFARVSGICVIE 425
Qy 305 RHHLHL-----LPQTSPPPKWNTSRSTRSASTQLLYDDHIGK----- 343
Db 426 DLCTSDVCVATSEVSVSLDPSKINSRELQQASEL-----DHVVELEKIAIVRLQOR 481
Qy 344 --VSLIGAGMRSHPGVTATFCEALAAVGVNIELIS 376
Db 482 AIISLIGNVEQSSLIEXT-GRVLRKSGVNVQMIS 515

RESULT 14
US-10-301-997-23
; Sequence 23, Application US/10301997

```

; APPLICANT: Falco, Saverio Carl
; APPLICANT: Famodu, Layo
; APPLICANT: Rafalski, Jan A.
; APPLICANT: Ramaker, Michael
; APPLICANT: Tarczyński, Mitchell C.
; APPLICANT: Thorpe, Catherine
; TITLE OF INVENTION: PLANT METHIONINE SYNTHASE GENE AND METHODS FOR INCREASING THE
; FILE REFERENCE: BB-1067-B
; CURRENT APPLICATION NUMBER: US/09/989,339
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 08/703,829
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-989-339-23

Query Match 14.2%; Score 310.5; DB 11; Length 449;
Best Local Similarity 30.2%; Pred. No. 4.6e-21;
Matches 102; Conservative 55; Mismatches 118; Indels 63; Gaps 8;

QY 1 VQYGGSSVADABRIRRVARIVATKQGNVVVVVSAMGDTTDDLLDLAQOVCP----- 55
Db 6 VSKFGTGVADFDAMRSADIVLS-----DANVLVLSAGITNLLVALAEGLEPGERFE 62
QY 56 -----APPPRELDMLLTAGERIS 73
Db 63 KLDAIRNIQFALLERLYPNVIREIERLENTVLAEAAALATSPALTDELVSHGELMS 122
QY 74 NALVAMAIESLGAHARSFTGSGAGVITTCGTHNAK-----IIDVTFGRLOTALBEGRVVL 128
Db 123 TLLFVILRERDVQAQWFDVRKY-METNDRFGRAEFPDIAALAEALQLLPRLNEG-LVI 180
QY 129 VAGFQGVSDTKDVTTLGRGSDTTAVAAALGADVCEIYTDVDGIFISADPRIYVNARK 188
Db 181 TQGFIG-SENKGETTLGRGSDYTHALLAEALHASRVDIWDVFGIYTTDPRVVSAAKR 239
QY 189 LDTVTEEMLEMAACGAKVLMRCVEYARRHNPVHRSSYSYSDRPGTVVVGSIKDVPMED 248
Db 240 IDEIAFAEAAENATFGAKVLHPATILPAVRSDIPVFGSSKDPKPRAGGTLVCNKTEP--- 296
QY 249 PILTGVADRSEAKVTIVGLPDI--PGYAAKVFAVAR 284
Db 297 EUFRALALRRNQTLTLHSLNMLHSRGLAEVFGILAR 334

Search completed: November 21, 2003, 16:38:11
Job time : 18.537 secs

QY 353 SHPGVTATFCEALAAVGNIELISTSEDQSRCCAAT 389
 |||||
 D β 356 SHPGVTATFCEALAAVGNIELISTSEIRISVLCDPT 392
 |||||

RESULT 2

G871199
aspartokinase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_charge 14-Dec-2001
C:Accession: G871199
R:Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; PMID:21128732; PMID:11234002
A:Accession: G871199
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STC>
A:Cross-references: GB:AL450380; NID:gl3093941; PIDN:CA031839.1; GSFDB:GN00147
C:Genetics:
A:Gene: ask
C:Superfamily: aspartate kinase; aspartate kinase homology

RESULT 3
S42422
N:alternate names: Mycobacterium smegmatis
C:Species: Mycobacterium smegmatis
C:date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 18-Jun-1999
C:Accession: S42422; S31801
R:Girillo, J.D.; Welsbrod, T.R.; Pascopella, L.; Bloom, B.R.; Jacobs Jr., W.R.
Mol. Microbiol. 11, 629-639, 1994
A:title: Isolation and characterization of the aspartokinase and aspartate semialdehyde
A:reference number: S42421; NUID:94254720; PMID:7910936
A:Accession: S42422
A:Molecule type: DNA

A;Residues: 1-421 <CIR>
A;Cross-references: EMBL:D17372; NID:g44506; PIDN:CAA78984.1.; PID:g581352
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
C;Genetics:
A;Start codon: GTG
C;Superfamily: aspartate kinase; aspartate kinase homology
C;Keywords: alternative initiators; phosphotransferase
F;1-421/Product: aspartate kinase alpha chain #status predicted <AS>
F;3-406/Domain: aspartate kinase homology <DKI>
F;250-421/Product: aspartate kinase beta chain #status predicted <ASB>

Query Match 67.4%; Score 1478.5; DB 2; Length 421;
Best Local Similarity 77.6%; Pred. No. 3.3e-102;
Matches 302; Conservative 31; Mismatches 35; Indels 21; Gaps 4;

QY 1 VKYGGSSVADAEIRRRVAERIVATKKOGNDVVVVVSAMGDTTDDLLDLAQVCPPAPPR 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5 VKYGGSSVADAEIRRRVAERIVETKKAAGNDVVVVVSAMGDTTDDLLDLARQVSPAPPR 64
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 EDMDMLLTAGERISNALVAWATESGAHARSFTGSOAGVITTTGHGNAKIIDVTPEGRJOTA 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 EDMDMLLTAGERISNALVAWATESGAQARSTGTGSOAGVITTTGHGNAKIIDVTPEGRJURDA 124
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 LEEGRVLVLVAGFGVQSQDPTKDVTTLGRGSGSTTAVAAAALGADVCEIYTDVDGIFSDP 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 125 LDQGQILVLVAGFGVQSQSDKVTVTLGRGSGSTTAVAAALDADVCBIYTDVDGIFTADP 184
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 RVRVARKLDTVTPEMLEMAACGAKVLMRCVEYARRHNIPVHRSSYSYDRPGTVVVG 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 185 RVVPVARHLDTVSFEEMLEMAACGAKVLMRCVEYARYRYNPVIHVRSSYSYDKPTIVKG 244
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 IKDVPMWEPIILTGVAFDRSEAKVTIVGLPDTPGYAAKFVRAVARRRQRHGAEE----- 295
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 245 IEDIWPEDALLTGVAHDRSEAKVTIVGLPDTPGVYA KFVRAVAE-----ADVNI DM 295
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 296 RIQQ-----RGQRDHHLHLPLQTSGGPPPKWNTRSETRSASTQLLDHHIGKVSLLGAG 350
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 296 VLQNISKIEDGKTD--ITFTCARDNGPRAVEKLALKSEIGFSQVLYDDHHIGKVSLLGAG 353
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 351 MRSHPGVATPFCEALAAAGVNIELISTSE 379
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 354 MRSHPGVATPFCEALAAEAGINIDLISTSE 382
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
S15276
aspartate kinase (EC 2.7.2.4) alpha chain - Corynebacterium glutamicum
N;Alternate names: aspartokinase alpha chain
C;Species: Corynebacterium Glutamicum
C;Date: 17-Apr-1993 #sequence revision 17-Apr-1993 #text_change 18-Jun-1999
C;Accession: S15276; S49977
R;Kalinowski, J.; Cremer, J.; Bachmann, B.; Eggeling, L.; Sahm, H.; Puehler, A.
Mol. Microbiol. 5, 1197-1204, 1991
A;Title: Genetic and biochemical analysis of the aspartokinase from Corynebacterium gluta
micum
A;Reference number: S15276; PMID:92065816; PMID:1956296
A;Accession: S15276
A;Molecule type: DNA
A;Residues: 1-421 <PUE>
A;Cross-references: EMBL:X57226; NID:g40509; PIDN:CAA40502.1; PID:g580983
A;Experimental source: ATCC 13032
R;Sererbriljski, I.; Wojcik, F.; Reyes, O.; Leblon, G.
submitted to the EMBL Data Library, November 1994
A;Description: Two loci of Corynebacterium glutamicum ATCC17965 that complement Escherichia coli strains lacking the ability to synthesize lysine
A;Reference number: S49977
A;Accession: S49977
A;Molecule type: DNA
A;Residues: 329-421 <GER>
A;Cross-references: EMBL:X82928; NID:g599716; PIDN:CAA58100.1; PID:g599717
R;Kalinowski, J.; Bachmann, B.; Thierbach, G.; Puehler, A.
Mol. Gen. Genet. 224, 317-324, 1990
A;Title: Aspartokinase genes lyc-alpha and lyc-beta overlap and are adjacent to the asp operon
A;Reference number: S12250; PMID:91094767; PMID:1980002
A;Contents: annotation

C;Genetics:
A;Gene: lycC-alpha
A;Start codon: GTG
C;Superfamily: aspartate kinase; aspartate kinase homology
C;Keywords: alternative initiators; phosphotransferase
F;3-406/Domain: aspartate kinase homology <DKI>

Query Match 56.1%; Score 1229.5; DB 2; Length 421;
Best Local Similarity 64.3%; Pred. No. 1e-83;
Matches 252; Conservative 55; Mismatches 58; Indels 27; Gaps 5;

QY 1 VOKYGGSSVADAERIRRVAAERIVATKKQGNDDVVVVVSAMGDTTDDLLDLAQVCPAPP 60
DB 5 VOKYGGSSLESAERIRNVAERIVATKKAGNDVVVVVSAMGDTTDELLEAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNAKIIVDTTPGRLOTA 120
DB 65 EMDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNAKIIVDTTPGRVREA 124

QY 121 LEEGRVVLVAGFGQVSDTKVITLGRGGSTTAVMAAALGADVCEIYDVGDFISADP 180
DB 125 LDEKICITVAGFGVNGKTRDVTTLGRGGSTTAVMAAALNADVCEIYSDVDGVYADP 184

QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSYSDRPGTVVVG 240
DB 195 RIVPNAQKLEKLSPEEMLELAAGSKILVLSVYARAFNVPVLRVSYSYNDPFTLIAGS 244

QY 241 IKDVPMEPIITGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
DB 245 MEDIPVEAVLTGVATDKSEAKVTIVGLISDRPGEAAKVRALADAENIDMVLQNVSSVE 304

QY 291 HGAERLQGRQD-RHLLHLLP-QTSPPPPKWNTRSETRSASTOLLYDDHIGKYSLLI 347
DB 305 DGTDTITTCPRSDGRRRAEILKKLVQV-----NW-----TNLYDDQVGRKSLV 350

QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 5
TJ35383
probable aspartokinase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 31-Jan-2000
C;Accession: TJ35383
R;Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A;Reference number: Z21576
A;Accession: TJ35383
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-425 <MUR>
A;Cross-references: EMBL:AL079348; PIDN: CAB45482.1; GSPDB: GN00070; SCOEDB: SC66T3.26C
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: ask; SCOEDB: SC66T3.26C
C;Superfamily: aspartate kinase; aspartate kinase homology

Query Match 52.0%; Score 1141; DB 2; Length 425;
Best Local Similarity 61.6%; Pred. No. 3.9e-77;
Matches 239; Conservative 54; Mismatches 79; Indels 16; Gaps 4;

QY 1 VOKYGGSSVADAERIRRVAAERIVATKKQGNDDVVVVVSAMGDTTDDLLDLAQVCPAPP 60
DB 5 VOKYGGSSVADAERIKRVAERIVATKKAGNDDVVVVVSAMGTTDELIDAEQVSPFAGR 64

QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNAKIIVDTTPGRLOTA 120
DB 65 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNAKIIVDTTPGRIRTS 124

QY 121 LEEGRVVLVAGFGQVSDTKVITLGRGGSTTAVMAAALGADVCEIYDVGDFISADP 180
DB 125 LEEGRVVLVAGFGQVSDTKVITLGRGGSTTAVMAAALGADVCEIYDVGDFISADP 180

DB 125 VDEGNVAIVAGFGVSDTKVITLGRGGSTTAVMAAALDADVCEIYDVGDFVFTADP 184
QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSYSDRPGTVVVG 240
DB 185 RIVPNAKKTIDWISFEDMLELAAGSKVLLHRCVEYARRVNIPIHVRSSFSGLQGTWVSSE 244
QY 241 -IK--DVPMEPIITGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARR-----RQHRH 291
DB 245 PIQGEKHVEQALISGVAHDTSEAKVTIVGVGPKFGEAAAFRAIADAGVNDIMVYQVNS 304
QY 292 GAAERLQGRQD-RHLLHLLP-QTSPPPPKWNTRSETRSASTOLLYDDHIGKYSLLIGAM 351
DB 305 ALSTGLTDS-----FTLPKSEGRKAIDALEKRPFGIFDSRLYDDQIGKISLVGAGM 357
QY 352 RSHPGVTATFCEALAAVGVNIELISTSE 379
DB 358 KSNPGVTADFTALSDAGVNIELISTSE 385

RESULT 6
C83531
aspartate kinase alpha and beta chain PA0904 [imported] - Pseudomonas aeruginosa (strain
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 14-Dec-2001
C;Accession: C83531
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brj
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: C83531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-412 <STO>
A;Cross-references: GB:AE004525; GB:AE004091; NID:G9946805; PIDN:AAG04293.1; GSPDB:GN001;
A;Experimental source: strain PA01
C;Genetics:
A;Gene: lycC; PA0904
C;Superfamily: aspartate kinase; aspartate kinase homology

Query Match 37.5%; Score 823; DB 2; Length 412;
Best Local Similarity 42.6%; Pred. No. 1.6e-53;
Matches 170; Conservative 80; Mismatches 105; Indels 44; Gaps 4;

QY 1 VOKYGGSSVADAERIRRVAAERIVATKKQGNDDVVVVVSAMGDTTDDLLDLAQVCPAPP 60
DB 5 VOKFGTSGVTVERIEQVAKVKKFEAGDDVVVVVSAMSGETNRNLIGLANQIMEQFVPR 64

QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNAKIIVDTTPGRLOTA 120
DB 65 ELDVYVSTGEQVTIALLSMALIKRGVPVSYTGQVRIILTDSATKARILHDDTHRAD 124

QY 121 LEEGRVVLVAGFGQVSDTKVITLGRGGSTTAVMAAALGADVCEIYDVGDFISADP 180
DB 125 LKAGRVVVVAGFGV-DGNGNITTLGRGGSDTTGVALAAALKADECIYTDVGVYTTDP 183

QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSYSDRPGTVVVG 240
DB 184 RVVPAQRLLDKITFEEMLEWASLGSKVLQIRAVEFAGKNVPLRVLHVSFQEGFGTILITD 243
QY 241 IKDVPMEPIITGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRQ----- 288
DB 244 DEESMEQPIISGIAFNDEAKLTIRGVPTPGVAFKILGPISAAVNEVDIMVQVNAHDN 303
QY 289 -----HEHGAERLQGRQDRHLLHLLPQTSPPPPKWNTRSETRSASTOLLYDDH 340
DB 304 TTDTFTVHKNDVYNAL-----ILKQTA-----ANIGAKEALGDTN 340
QY 341 IGKVSLLIGAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
DB 341 IAKVSIVGVGMRSHAGVASRMFEALAKESINIQMISTSE 379

RESULT 7

S76764
hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
C/Species: *Synechocystis* sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C/Accession: S76764
R/Kaneko, I.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shampo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
s.
A/Reference number: S74322; MUID:97061201; PMID:8505231
A/Accession: S76764
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-600 <KAN>
A/Cross-references: EMBL:D90916; GB:AB001339; NID:g1653715; PIDN:BAAL8676.1; PID:d101940
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C/Superfamily: aspartate kinase homology
F:3-417/Domain: aspartate kinase homology <DKI>

Query Match 37.2%; Score 816.5; DB 2; Length 600;
Best Local Similarity 43.9%; Pred. No. 7.6e-53;
Matches 183; Conservative 83; Mismatches 118; Indels 33; Gaps 7;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDDLAQVCPAPPR 60
DB 5 VQRFSGTSGTGTBRISQAVQRIKRTVQGGNSLVVVVSAMGKSTDLVLDIAQISNPORR 64
QY 61 ELDMLTAGERISNALVAVAIESLGAHARSFTGSGQAGVITTTGHNAKIIDVTPGRLQTA 120
DB 65 EMDMLSTGEQVSIALLSLALQIDQPAISLTGAQGVIVTEAHSRRARILEIRPDRLEHH 124
QY 121 LEEGRVLVAGFGVSGQ-DTKQVTLTGRGSDTTAVAMAALGADVCEIYTDVDFISAD 179
DB 125 LREGKVVVAGFGTSSVSEHLEITTLGRGSDTSAVALAAALKADFCBIYTDVPGILTTD 184
QY 180 PRIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVG 239
DB 185 PRLVPEQLWAETTCDEMLELALSLGAKVLPRAVEIARNYGIPLVRSWSDEPGTKVA 244
QY 240 SIKDVMEDPILT-----GVADHREAKVTIVGLDPDIFYAAKVFRAVARR----- 286
DB 245 P-----PVQNRSLVGLSRIAKAVDGVEDQAKVALLRVPDRPGVASKLFRDIAQQQVDD 300
QY 287 ---RQHRGGAERLQGRQDRHLLLPQTSGPPPKWTRSETRSASTOLLYDDHIGK 343
DB 301 LIQTHDGNNDNIATVVKDLNLTAEVTSIAFALRSYPEADQEA---EIIIVEKGIAK 357
QY 344 VSLIGAGMESHGPGVATFCEALAAVGVNIELISTSE-----DQR--SRCAATPN 391
DB 358 IATAGAGMIGRPGIAAKMFKTLADVGVIEMISTSEVKVSCVIDQORDADRAIALSN 414

RESULT 8

A48946
aspartate kinase (EC 2.7.2.4) II precursor - *Bacillus* sp. (strain MGA3)
N/Alternate names: lysine-sensitive aspartokinase II
N/Contains: aspartokinase II alpha chain; aspartokinase II beta chain
C/Species: *Bacillus* sp.
C/Date: 21-Jan-1994 #sequence_revision 24-Feb-1995 #text_change 18-Jun-1999
C/Accession: A48946; B48946; C48946
R/Schudel, F.J.; Flickinger, M.C.
Appl. Environ. Microbiol. 58, 2806-2814, 1992
A/Title: Cloning and nucleotide sequence of the gene coding for aspartokinase II from a
A/Reference number: A48946; MUID:93073880; PMID:1444390
A/Accession: A48946
A/Molecule type: DNA
A/Residues: 1-411 <SCH>
A/Cross-references: GB:M93419; NID:g142539; PIDN:AAA22251.1; PID:g142540
A/Experimental source: strain MGA3

A/Note: sequence modified after extraction from NCBI backbone
A/Note: sequence extracted from NCBI backbone (NCBIN:117802)

A/Accession: B48946
A/Molecule type: protein
A/Residues: 2-9 <SC2>
A/Note: this material was purified alpha chain
A/Accession: C48946
A/Molecule type: protein
A/Residues: 246-253 <SC3>
A/Note: this material was purified beta chain; the beta chain apparently was not blocked
C/Comment: The amino terminal Met of the beta chain is preceded by a strongly predicted
odon was demonstrated after deletion of the alpha chain region. This suggests that expres
C/Genetics:
A/Start codon: GTG
C/Superfamily: aspartate kinase; aspartate kinase homology
F:2-245/Product: aspartokinase II alpha chain #status predicted <ALH>
F:3-401/Domain: aspartate kinase homology <DKI>
F:246-411/Product: aspartokinase II beta chain #status predicted <BET>

Query Match 35.8%; Score 785; DB 2; Length 411;
Best Local Similarity 42.1%; Pred. No. 1e-50;
Matches 170; Conservative 81; Mismatches 97; Indels 56; Gaps 6;

QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDDLAQVCPAPPR 60
DB 5 VQRFSGTSGTGTBRISQAVQRIKRTVQGGNSLVVVVSAMGKTTDELVDLAKQISAHPPKR 64
QY 61 ELDMLTAGERISNALVAVAIESLGAHARSFTGSGQAGVITTTGHNAKIIDVTPGRLQTA 120
DB 65 EMDMLTGTGQVTSISLLAMALNEKYEATSYTQWQAGITTEPVFGNARILNTEIKIQK 124
QY 121 LEEGRVLVAGFGVSGQ-DTKVTLTGRGSDTTAVAMAALGADVCEIYTDVDFISADP 180
DB 125 LNEGKVVVAGFGQIDEH-GEITTLGSGSDTTAVALAALAAKAEKCDIYTDVGTVTDP 183
QY 181 RIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVG 240
DB 184 RYVKSARKLASISYDEMLELANLGAGVLPRAVEFARVNGITILEVRSSMEREBEGTIE-- 241
QY 241 IKQVPEM-DPIILGVADHRESEAKVTIVGLDPDIFYAAKVFRAVARRRQHRHGAERLQ 299
DB 242 -EVTMEQNJVVRGVAFEDBITRTVFGLPNSLTSITFTLLA----- 284
QY 300 RGRQDRHLLLPQTSGPPPKWTRSETRSASTOLLYDD----- 339
DB 285 ---QNRINVDIIQSA-----TDAETTNLSFSIKSDDEETMAVLENNKLLNYQGI 333
QY 340 ---HICKVSLIGAGMESHGPGVATFCEALAAVGVNIELISTSE 379
DB 334 ESETGLAKAVSIGVSGMISNPGVAAKMFVLAALNGIQVKWVSTSE 377

RESULT 9

JC4640
aspartate kinase (EC 2.7.2.4) II - *Bacillus* stearothermophilus
N/Alternate names: aspartokinase II
C/Species: *Bacillus* stearothermophilus
C/Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 18-Jun-1999
C/Accession: JC4640
R/Cantoni, R.; Labo, M.; De Rossi, E.; Riccardi, G.
Gene 169, 135-136, 1996
A/Title: Sequence of the *Bacillus* stearothermophilus gene encoding aspartokinase II.
A/Reference number: JC4640; MUID:96186919; PMID:8635739
A/Accession: JC4640
A/Molecule type: DNA
A/Residues: 1-407 <CAN>
A/Cross-references: GB:L46351; NID:g928810; PIDN:AAB06216.1; PID:g928811
C/Comment: This enzyme is the first key enzyme in the biosynthetic pathway of the aspartate
C/Genetics:
A/Gene: ask
C/Superfamily: aspartate kinase; aspartate kinase homology
C/Keywords: aspartate proteinase; phosphotransferase

F3-397/Domain: aspartate kinase homology <DKI>

Query Match 35.4%; Score 778; DB 2; Length 407;
Best Local Similarity 43.2%; Pred. No. 3.4e-50;
Matches 173; Conservative 77; Mismatches 98; Indels 52; Gaps 7;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKQGNDDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
Db 5 VQKFGGTSVGSIERIQHVANRVI BEVQKGNDDVVVVVVSAMGKTTDELVNLAQKISNHPSKR 64

Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLQTA 120
Db 65 EYDMLLTSTGEQVITALLSNALQELQPAISMVTAQVIGVTEAEHTARILHETERVEGH 124

Qy 121 LEEGRVVLVAGFGQVGSQD-TKDVTLTGRGSDTTAVAMAALGADVCEIYTDVGDIFPADP 180
Db 125 LDEGAIVTVAGFGVET-ETGEBITTLGRGSDTTAVAAALKAEEKCDIYTDVGVFTTDP 183

Qy 181 RIVRNARKLDTVTTFEEMLEVAACGAKVLMRCVYARRHNPVHVRSSYSYSDRPGTVVVG 240
Db 184 RYVTKARKIKSISYDEMLANLGAHVLPRAVEFAKNYEVLPVRSMSMENERGTWKV-- 241

Qy 241 IKDVPMEDEP-ILTVGAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRRQHRHGAERLOG 299
Db 242 -EEVSMEOHLIVRGIAFEDQVTRVTVGIEKVLQSVATIFTALNR----- 286

Qy 300 RGRQRHLLHLLPQTSGLPPPKWNTSRSTRAS-----TOLL-----YDD 339
Db 287 -----GINVDIIQ-----NATNSETASVSIRTEDLPETLQVLQALEGADVHYES 333

Qy 340 HIGKVSLLTGAGMRSHPGVTATFCBALAAGVNIELISTSE 379
Db 334 GLAKVSIIVGSMINPGVAARVFEVLADQGIKXVSISE 373

RESULT 10

AE2261

C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AE2261
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, R.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2261
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-606 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075343.1; PID:gi17132777; GSPDB:GN00179
A:Experimental source: strain PCC 7120

C:Genetics:
A:Gene: alr3644

Query Match 35.4%; Score 776.5; DB 2; Length 606;
Best Local Similarity 43.2%; Pred. No. 7.2e-50;
Matches 175; Conservative 74; Mismatches 115; Indels 41; Gaps 7;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKQGNDDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
Db 5 VQKFGGTSVGSVERIQAVQRYKTVKAGNSLVVVVVSAMGKTTDGLVLANEISRNPNRR 64

Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLQTA 120
Db 65 EYDMLLTSTGEQVITALLSNALQELQPAISMVTAQVIGVTEAEHTARILHETERVEGH 124

Qy 121 LEEGRVVLVAGFGQVGSQD-TKDVTLTGRGSDTTAVAMAALGADVCEIYTDVGDIFPAD 179
Db 125 IXEKGKVVVAGFGQISNGSLEITTLGRGSDTSAVALAALQADFCEIYTDVPGILATD 184

Qy 180 PRIVRNARKLDTVTTFEEMLEVAACGAKVLMRCVYARRHNPVHVRSSYSYSDRPGTVV 238

Db 185 PRIVPEAQLMGITTCNEMLELASLGAHVLPRAVEIARNYGMPLVVKSWTDDPGTWT 244

Qy 239 -----GSKIDVPMEDPILTVGAHDRSEAKVTIVGLPDIPGYAAKVFRAVARR-----R 287
Db 245 XPQGRSLINLELAREV-DDVELDQAKVALLRPDKPGVAAKLFGEISRCKVDVDLII 303

Qy 288 QHRHGAERLOGR-----RODRHLLHLLPQTSGLPPPKWNTSRSTRASQ 334
Db 304 QSIH-----EGNSNDIAFTVTTPILKRAEVAVAIAFSLRSP-----SHPKSDAE 349

Qy 335 LLYDDHIGKVSLLTGAGMRSHPGVTATFCBALAAGVNIELISTSE 379
Db 350 VMVEQNIKVSITAGAGMIGRPGVAAQMFATLAEAGVNIQMISTSE 394

RESULT 11

H81865

C:Species: Neisseria meningitidis
A:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: H81865
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell, S.; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: H81865
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-405 <PAR>
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84929.1; PID:g738034;
A:Experimental source: serogroup A, strain Z2491

C:Genetics:
A:Gene: lygC; NMA1701
C:Superfamily: aspartate kinase; aspartate kinase homology
C:Keywords: phosphotransferase

Query Match 34.7%; Score 761; DB 2; Length 405;
Best Local Similarity 40.8%; Pred. No. 6.1e-49;
Matches 161; Conservative 82; Mismatches 116; Indels 36; Gaps 3;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKQGNDDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
Db 5 VHYGTSVGSVERIKNKAARVAKAEGHDI VVVVVSAMSGETNRLVALAHEHQPDP 64

Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLQTA 120
Db 65 ELDVLTSTGEQVITALLSNALQELQPAISMVTAQVIGVTEAEHTARILHETERVEGH 124

Qy 121 LEEGRVVLVAGFGQVGSQD-TKDVTLTGRGSDTTAVAMAALGADVCEIYTDVGDIFPADP 180
Db 125 LTAGKVVIVAGFGQISSE-GD1STLGRGSDTSAVALAALKADEQIYTDVGVYTTDP 183

Qy 181 RIVRNARKLDTVTTFEEMLEVAACGAKVLMRCVYARRHNPVHVRSSYSYSDRPGTVVVG 240
Db 184 RVVPEARMOTVTTFEEMLEVASLGSVQLQSRVFEAGKYKVRVLRVLSLQGGNGILITF 243

Qy 241 IKDVPMEDEPILTVGAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRRQHRHGAERLOGR 300
Db 244 EEDDNMERAATVGIAPDKNQARINVRGVPDKPGVAYQILGAVADANIE----- 291

Qy 301 GRQRHLLHLLPQTSGLPPPKWNTSRSTRASQTLV-----DDHLGKV 344
Db 292 -----VDMIIQNVGSEGTTFDFVTPRGDYKQTEILSERQDSIGAASIDGDDTVCKV 344

Qy 345 SLIGAGMRSHPGVTATFCBALAAGVNIELISTSE 379
Db 345 SAVGLGRSHVGVAAKIFRLAEBGINIQMISTSE 379

RESULT 12

F81076

aspartokinase, alpha and beta chains NMB1498 [imported] - *Neisseria meningitidis* (strain C/Species: *Neisseria meningitidis*)
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C/Accession: F81076
 R/Tetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.; Science 287, 1809-1815, 2000
 A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, B.R.; Rappuoli, R.; Vitti, A.; Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
 A/Reference number: A81000; MUID:20175755; PMID:10710307
 A/Accession: F81076
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-405 <TET>
 A/Cross-references: GB:AE002499; GB:AE002098; NID:G7226737; PIDN:AAF41854.1; PID:G722674
 A/Experimental source: serogroup B, strain MC58
 C/Genetics:
 A/Gene: NMB1498
 C/Superfamily: aspartate kinase; aspartate kinase homology

Query Match 34.7%; Score 760; DB 2; Length 405;
 Best Local Similarity 40.8%; Pred. No. 7.3e-49;
 Matches 161; Conservative 81; Mismatches 117; Indels 36; Gaps 3;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 Db 5 VHYGGTSVGSPEIRKNVAKVAKAABGHDIIVVVSAMSGETNRLVALAHQEHDPDR 64
 Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIDVTPGRLOTA 120
 Db 65 ELDVLTAGEQVITGLLAWALKDIGVDAKSYTGQVALXTDTAHTKARIESIDDEKWRAD 124
 Qy 121 LEEGRVVLVAGFGVSGDQKDVTTGLRGSGDITAVAMAALGADVCEIYTDVDFGSADP 180
 Db 125 LTAGKVVIVAGFGIASE-GDITLGRGSDTSVALAALKADECGIYTDVDFGVITDP 183
 Qy 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRGTVVWGS 240
 Db 184 RVVPEARRMDTVTPEEMLESLGSKVLQIRSVFAGKVKVLRVLSSLDQGGNGTLITF 243
 Qy 241 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDI PGYAAKVFRAVARRRRQHRHGAERLQGR 300
 Db 244 EEDNNVERAAVTGIAFDKQARINVRGPKGVAQILGAVADANIE----- 291
 Qy 301 GQDRHHLHLPOTSGPPPKWTRSETRSASTQLLY-----DDHIGKV 344
 Db 292 -----VDMIIQNVGSEGTDFSFVPRGDYKQTLLEILSERQDSIGAASIDGDDTVCKV 344
 Qy 345 SLIGAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 Db 345 SAVGLGMRSHVGVAAKIFRTLAEBEGINIQMISTSE 379

RESULT 13

H84036

aspartokinase II alpha and beta subunits lysC [imported] - *Bacillus halodurans* (strain C/Species: *Bacillus halodurans*)
 C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C/Accession: H84036
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Fujii, F.; Hiraoka, T.; Nucleic Acids Res. 28, 4317-4331, 2000
 A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A/Reference number: A83650; MUID:20512562; PMID:11058132
 A/Accession: H84036
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-412 <STO>
 A/Cross-references: GB:AP001517; GB:BA000004; NID:G10175500; PIDN:BA066815.1; GSPDB:GN00
 A/Experimental source: strain C-125
 C/Genetics:
 A/Gene: lysC
 C/Superfamily: aspartate kinase; aspartate kinase homology

Query Match 34.5%; Score 757.5; DB 2; Length 412;
 Best Local Similarity 42.0%; Pred. No. 1.1e-48;
 Matches 173; Conservative 64; Mismatches 102; Indels 73; Gaps 7;
 Qy 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 Db 5 VQYGGTSVGSVQRIEIVANRIIATAASGKXVVVVSAMGKTTDELVLGLASEITDEPRKR 64
 Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIDVTPGRLOTA 120
 Db 65 EMDMLLTAGEQVITALLAMALHKKGEQATSLTGWAGNITESHGHNARIIEVERARDEL 124
 Qy 121 LEEGRVVLVAGFGVSGDQKDVTTGLRGSGDITAVAMAALGADVCEIYTDVDFGSADP 180
 Db 125 LDKVHVIVIVAGFGIDQO-GBIMTGLRGSGDTSVALAALKAEBCIYITDVTGVFTDP 183
 Qy 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRGTVVWGS 240
 Db 184 RVVQARLQITISYDEMLEMANIGASVHLHPRAVEFAKNIYIPLLVASSWVDEPGTME-- 241
 Qy 241 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDI PGYAAKVFRAVARRRRQHRHGAERLQGR 299
 Db 242 -EEGTMEQNLLVRGIAFENDVTKVTHGLP-----QT-SGPPPKWTRSETRSASTQLL 336
 Qy 300 RGRDRHHLHLLP-----YDDHIGKVLICAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 Db 271 -----NEIHTLPTLFAHLSAAGINVDIIIONQDTSMAISFISITQSSSLKETLARLEELK 324
 Qy 337 -----YDDHIGKVLICAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 Db 325 GELSFQGDIEEELAKVISVSGMVSNPFGVAAKMFELANDIAIKMVSTSE 376

RESULT 14

E81405

aspartate kinase (EC 2.7.2.4) alpha and beta chains Cj0582 [imported] - *Campylobacter jejuni* C/Species: *Campylobacter jejuni*
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C/Accession: E81405
 R/Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrall Nature 403, 665-668, 2000
 A/Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals type A/Reference number: A81250; MUID:20150912; PMID:10688204
 A/Accession: E81405
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-400 <PAR>
 A/Cross-references: GB:AL139075; GB:AL111168; NID:G6967817; PIDN:CAB75218.1; PID:G6968049
 A/Experimental source: serotype O2, strain NCTC 11168
 C/Genetics:
 A/Gene: lysC; Cj0582
 C/Superfamily: aspartate kinase; aspartate kinase homology
 C/Keywords: phosphotransferase

Query Match 33.9%; Score 742.5; DB 2; Length 400;
 Best Local Similarity 39.9%; Pred. No. 1.4e-47;
 Matches 157; Conservative 89; Mismatches 110; Indels 37; Gaps 4;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 Db 4 VQYGGTSVGTLEIEIAVANRVIQSAQGNQLVVVVSAMSGVNTLIEQAEYFSKTPNGK 63
 Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIDVTPGRLOTA 120
 Db 64 DMDMLSSGERVTSALLALNEKGPALSFSGRKAGIITDSVFTKARLHHTDKAIKSE 123
 Qy 121 LEEGRVVLVAGFGVSGDQKDVTTGLRGSGDITAVAMAALGADVCEIYTDVDFGSADP 180
 Db 124 LONGKIVIVAGFGVDEE-GNVTTLGRGSDLSAVAGALNADLCEIYTDVDFGVITDP 182
 Qy 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRGTVVWGS 240

```
Db      183 RIEPKAKKLDKISYEEMLESLGAKVLQNRSEVELAKKLNVLVTRSSFNNNEGTMIT-- 240
Qy      241 IKDVPMDPIITGVADHRSKAKVTIVGLDPIPGYAAKVFRVARRRRQHRHGAERLQGR 300
Db      241 -KEDMEQALVSGIALDKQARVTRNIEDKPGIAAEIFSVLA----- 282
Qy      301 GRDRHHLHLLPQTSQPPPKQW-----TRSETRSASTOLLYDDHIGKVS 346
Db      283 --NENINVDMIQNVGDGATNLGFTVPQNELELAKVAMQKILSSKTTIESDSAVLKVSI 340
Qy      347 IGAGMRSHPGVATFCBALAAVGVNIELISTSE 379
Db      341 VGVGKSHSGVASKAFKALADEGINIGMISTSE 373

RESULT 15
B70399
aspartokinase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 18-Jun-1999
C:Accession: B70399
E:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: B70399
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-415 <AQF>
A:Cross-references: GB:AE000726; NID:g2983612; PIDN:AAC07182.1; PID:g2983613; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: lycC
C:Superfamily: aspartate kinase; aspartate kinase homology
E:3-400/Domain: aspartate kinase homology <DKI>

Query Match      33.5%; Score 734.5; DB 2; Length 415;
Best Local Similarity 42.4%; Pred. No. 5.9e-47;
Matches 164; Conservative 76; Mismatches 124; Indels 23; Gaps 6;

Qy      1 VQKGGSSVDAEIRIRVAERIVATKQGNDDVVVVVSAWGDTDDLLDLAQOVCPPPPR 60
Db      5 VQKGGTSVGTLEIEICAKKVIKALEKGYKPVVVSAMAGETDSLIELAKKVHPPR 64
Qy      61 ELDMLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTPGRLOTA 120
Db      65 EMDTLLAVGEQKATLPAWTLNKLGYPAVSLCGWQVPIITDNVHTKARIKIGISRLSL 124
Qy      121 LEEGRVVLVAGFGVSGDTKDVTLGRGSGDTTAVMAAALGADVCEIYTDVDPISADP 180
Db      125 IEEGYIPVAGFGVGTEDW-EITLGRGSGDTTAVALLAALKAD-CBINTDVPGVFTADP 182
Qy      181 RIVNARKLDVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSYSDRPGTVVVG 240
Db      183 RIVNARKIKKISYEEMLESLGAKVQIRSEVFAKYNVIRVSTFEEEGTWIVP- 241
Qy      241 IKDVPMDPIITGVADHRSKAKVTIVGLDPIPGYAAKVFRVARRRRQHRHGA----- 294
Db      242 -EDEVMEKVAVRGITVDTKETRFTVVRVPDKPGIAABIFKAL-----GDAHIVVDM 291
Qy      295 --ERLQGRGRDRHHLHLLPQTSQPPPKWTRSETSSASTQLLYDDHIGKVSILGAGR 352
Db      292 IVQNVSHGEGYDMS--FTVNTDADKAEIIVKIAQKIGAEVVRDDKAKISIVGIGMK 349
Qy      353 SHPGVATFCBALAAVGVNIELISTSE 379
Db      350 SAYGVAGRMFDILARNGINIKAIISTSE 376
```

Search completed: November 21, 2003, 16:09:51
Job time : 9.84535 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 4.9955 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-4
Perfect score: 2193
Sequence: 1 VQYGGSSVADAEIRRVAE.....SAATRRPCTRGDRWACQ 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match %	Length	DB ID	Description
1	1615.5	73.7	421	1	AK_MYCTU	P97048 mycobacteri
2	1478.5	67.4	421	1	AK_MYCSM	P41403 mycobacteri
3	1248.5	56.9	421	1	AK_COREF	Q8rqnl corynebacte
4	1224.5	55.8	421	1	AK_CORGL	P26512 corynebacte
5	1216.5	55.5	421	1	AK_CORFL	P41398 corynebacte
6	823	37.5	412	1	AK_PSEAE	O69077 pseudomonas
7	819	37.3	405	1	AK_THERH	P97151 thermus the
8	785	35.8	411	1	AK2_BACSP	O59228 bacillus sp
9	778	35.5	407	1	AK2_BACST	P53553 bacillus st
10	734.5	33.5	415	1	AK_AQUAE	O67221 aquifex aeo
11	721	32.9	405	1	AK_HELPJ	Q9zjz7 helicobacte
12	712.5	32.5	408	1	AK2_BACSU	P08495 bacillus su
13	711	32.4	405	1	AK_HELPV	O25827 helicobacte
14	532	24.3	404	1	AK1_BACSU	O04795 bacillus su
15	492	22.4	473	1	AK1_METJA	O57931 methanococc
16	480.5	21.9	446	1	AK1_RICPR	Q9zc17 rickettsia
17	364.5	16.6	819	1	AK1H_SERMA	P27725 serrattia ma
18	359.5	16.4	820	1	AK1H_ECOLI	P00561 escherichia
19	314.5	14.3	815	1	AKH_HAEIN	P44505 haemophilus
20	311.5	14.2	921	1	AKH_DAUCA	P37142 daucus caro
21	310.5	14.2	449	1	AK3_ECOLI	P08660 escherichia
22	306.5	14.0	454	1	AK3_BACSU	P94417 bacillus su
23	284	13.0	816	1	AKH_BUCAI	P57290 buchnera ap
24	280	12.8	431	1	AK1_CHLTR	O84367 chlamydia t
25	270	12.3	920	1	AKH1_MAIZE	P49079 zea mays (m
26	269	12.3	437	1	AK1_CHLMU	O9PK32 chlamydia m
27	266	12.1	917	1	AKH2_MAIZE	P49080 zea mays (m
28	264	12.0	814	1	AKH_EUCAP	Q8K9U9 buchnera ap
29	241	11.0	440	1	AK1_CHLPN	Q92610 chlamydia p
30	238	10.9	125	1	AK1_MYCBO	P47731 mycobacteri
31	227.5	10.4	527	1	AK1_YEAST	P10869 saccharomyc
32	218	9.9	809	1	AK2H_ECOLI	P00562 escherichia
33	210	9.6	519	1	AK_SCHPO	O60163 schizosacch

ALIGNMENTS

RESULT 1

AK_MYCTU	STANDARD;	PRT;	421 AA.
AC	P97048: O69676; P97181;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Aspartokinase (EC 2.7.2.4) (Aspartate Kinase)		
DE	alpha subunit (ASK-alpha); Aspartokinase beta subunit (ASK-beta).		
GN	ASK OR RV3709C OR MT3812 OR MTV025.057C.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37Rv;		
RA	Gilker J.M., Jucker M.T.;		
RT	"Mycobacterium tuberculosis ask-alpha, ask-beta and asd genes.";		
RL	Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37Rv;		
RC	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,		
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,		
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,		
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,		
RA	Sulston J.E., Taylor K., Whitehead S., Barrell S.G.;		
RT	"Deciphering the biology of Mycobacterium tuberculosis from the		
RT	complete genome sequence.";		
RL	Nature 393:537-544 (1998).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CDC 1551 / Oshkosh;		
RC	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,		
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,		
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,		
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,		
RA	Bishai W.;		
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and		
RT	laboratory strains.";		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
CC	-!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-		
CC	aspartate.		
CC	-!- ENZYME REGULATION: FEEDBACK INHIBITION BY LYSINE AND THREONINE (BY		
CC	SIMILARITY).		
CC	-!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING		
CC	FROM ASP TO THE CELL WALL PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,		
CC	TO MET, TO ILE AND TO THR.		
CC	-!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2		
CC	isoforms Beta (function not known) (By similarity).		
CC	-!- ALTERNATIVE PRODUCTS:		

34 175 8.0 306 1 ARGB_STRCO Q11a3 streptomyc

35 174.5 8.0 239 1 PYRH_BARBA Q8rt65 bartonella

36 151 6.9 226 1 PYRH_METJA Q85656 methanococc

37 150 6.8 239 1 PYRH_FUSNN Q8r695 fusobacteri

38 149 6.8 318 1 ARGB_BIFLO P59295 bifidobacte

39 146 6.7 241 1 PYRH_AGRTS Q8ufm1 agrobacteri

40 145.5 6.6 242 1 PYRH_ANASP Q8vxx5 anabaena sp

41 145.5 6.6 242 1 PYRH_ZYMMO Q8x5e9 zymomonas m

42 145 6.6 302 1 ARGB_STRCL Q9ic56 streptomyc

43 143.5 6.5 240 1 PYRH_XANCP P59009 xanthomonas

44 143 6.5 294 1 ARGB_MYCTU P94989 mycobacteri

45 141.5 6.5 240 1 PYRH_XANAC P59008 xanthomonas


```
Qy 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLLDLAQVCPAPPR 60
Db 5 VQYGGSSVADAEIRIRVAERIVETKKAGNDVVVVSAMGDTTDDLLDLARQVSPAPPR 64
Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGHNAKIIDVTPGRLOTA 120
Db 65 EMDMLLTAGERISNALVAMAEISLGAARSFTGSQAGVITTTGHNAKIIDVTPGRLRDA 124
Qy 121 LEEGRVVLVAGFGQVSGODTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 180
Db 125 LDEGQIVLVAGFGQVSGQSDKDVTTLGRGSGDTTAVAAALDADVCEIYTDVDFISADP 184
Qy 181 RIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSYDRGTVVGS 240
Db 185 RIVNARHLDTVSFEEMLEMAACGAKVLMRCVEYARRYNPVHVRSSYSYDKFTIVKGS 244
Qy 241 IKDVPMDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFAVARRRQHRHGAAE----- 295
Db 245 IEDIPMEDAILTGVAHDRSEAKVTIVGLPDIPGYAAKVFAVAE-----ADVNDIM 295
Qy 296 RLQG-----RGRQRHLLHLLPQTSQPPPKWTRSETRSASTQLLYDDHIGKVSILGAG 350
Db 296 VLQNIKIEBKTD--ITFTCARDNGPRAVEKLSALKSEIGFQVLYDDHIGKVSILGAG 353
Qy 351 MRSHPGVATFCEALAAVGVNIELISTSE 379
Db 354 MRSHPGVATFCEALAEAGINIDLISTSE 382
RESULT 3
AK COREP
ID AK COREP STANDARD; PRT; 421 AA.
AC OAEQNL;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
DE alpha subunit; Aspartokinase beta subunit].
GN LYSC OR ASK OR CE0220.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Itaya H., Kimura E., Kawahara Y., Hino Y., Kikuchi H., Nakamura Y.,
RA Kawaiabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeno K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- ENZYME REGULATION: FEEDBACK INHIBITION BY LYSINE AND THREONINE.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO THE CELL WALL PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,
CC TO MET, TO ILE AND TO THR.
CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
CC isoforms Beta (function not known).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative initiation;
CC Comment-2 isoforms, Alpha/Aspartokinase alpha subunit (shown
CC here) and Beta/Aspartokinase beta subunit, are produced by
CC alternative initiation;
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB081129; BAB88820.1; -; ALT_INIT.
DR EMBL; AF005214; BAC17030.1; -; ALT_INIT.
DR InterPro; IPR001048; AA_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aak_kinase; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRfam; TIGR00656; asp_kin_monofn; 1.
DR TIGRfam; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Alternative initiation; Complete proteome.
FT CHAIN 1 421 ASPARTOKINASE, ISOFORM ALPHA.
FT CHAIN 250 421 ASPARTOKINASE, ISOFORM BETA.
FT INIT_MET 250 250 FOR ISOFORM BETA.
SQ SEQUENCE 421 AA; 44793 MW; F0D638306983CB5F CRC64;
Query Match 56.9%; Score 1248.5; DB 1; Length 421;
Best Local Similarity 65.2%; Pred. No. 1.7e-80;
Matches 251; Conservative 58; Mismatches 63; Indels 13; Gaps 3;
Qy 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLLDLAQVCPAPPR 60
Db 5 VQYGGSSLESARIRNVAERIVATKKAGNDVVVVCSSAMGDTTDELLDLAAVNPVPPAR 64
Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGHNAKIIDVTPGRLOTA 120
Db 65 EMDMLLTAGERISNALVAMAEISLGAARSFTGSQAGVITTTGHNARIVDTPGRVREA 124
Qy 121 LEEGRVVLVAGFGQVSGQDKVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 180
Db 125 LDEGKICIVAGFGQVGNKETRDVTTLGRGSGDTTAVALLAALGADVCEIYSDVGYTADP 184
Qy 181 RIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSYDRGTVVGS 240
Db 185 RIVNARQLDTERLSFEEMLEMAAVGSKILVRSVEYARRNVPVRVSSVSNDFGLIAS 244
Qy 241 IKDVPMDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFAVARRR-----RQHRHGAA 294
Db 245 MEDIPWEBAVLGTGATDKSEAKVTVLGI PDKCEAAKVFRALADAEINIDMVLVNVSSVE 304
Qy 295 ERLQGRGRDRHLLHLLPQTSQPPPKWTRSETRSASTQLLYDDHIGKVSILGAGMRSH 354
Db 305 D-----GTTD--ITFTCPRSDDGPRAVELLKKMQQGDWNTNLYDDQVGVSLVAGMKSH 357
Qy 355 PGVTATFCEALAAVGVNIELISTSE 379
Db 358 PGVTAEFMEALRDVNVNVELISTSE 382
RESULT 4
AK CORGL
ID AK CORGL STANDARD; PRT; 421 AA.
AC P26512; Q59286;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
DE alpha subunit; Aspartokinase beta subunit].
GN LYSC OR CGL0251.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
```

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RX MEDLINE=92065816; PubMed=1956296;
 RA Kalinowski J., Cremer J., Bachmann B., Eggeling L., Sahm H.,
 RA Puehler A.;
 RT "Genetic and biochemical analysis of the aspartokinase from
 RT Corynebacterium glutamicum.";
 RL Mol. Microbiol. 5:1197-1204 (1991).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RX MEDLINE=94161495; PubMed=8117072;
 RA Patek M., Krumbach K., Eggeling L., Sahm H.;
 RT "Leucine synthesis in Corynebacterium glutamicum: enzyme activities,
 RT structure of leuA, and effect of leuA inactivation on lysine
 RT synthesis.";
 RL Appl. Environ. Microbiol. 60:133-140 (1994).
 RN [4]
 RN SEQUENCE OF 1-51 FROM N.A.
 RP STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RX MEDLINE=91094767; PubMed=1980002;
 RA Kalinowski J., Bachmann B., Thierbach G., Puehler A.;
 RT "Aspartokinase genes lysC alpha and lysC beta overlap and are
 RT adjacent to the aspartate beta-semialdehyde dehydrogenase gene asd in
 RT Corynebacterium glutamicum.";
 RL Mol. Genet. 224:317-324 (1990).
 CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
 CC aspartate.
 CC -!- ENZYME REGULATION: FEEDBACK INHIBITION BY LYSINE AND THREONINE.
 CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
 CC FROM ASP TO THE CELL WALL. PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,
 CC TO MET, TO ILE AND TO THR.
 CC -!- SUBUNIT: Tetramer consisting of two isoforms Alpha (catalytic) and
 CC two isoforms Beta (function not known).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative initiation;
 CC Comment=2 isoforms, Alpha/Aspartokinase alpha subunit (shown
 CC here) and Beta/Aspartokinase beta subunit, are produced by
 CC alternative initiation;
 CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X57226; CAA40502.1; -
 DR EMBL; X57226; CAA40503.1; -
 DR EMBL; AF005274; BAB97644.1; -
 DR EMBL; X70959; CAA50296.1; ALT_SEQ.
 DR PIR; I40723; I40723
 DR PIR; S15276; S15276.
 DR InterPro; IPR001048; Aa_kinase.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR005260; Asp_kin_monofn.
 DR InterPro; IPR001341; Aspartate_kinase.
 DR Pfam; PF00696; aak kinase; 1.
 DR Pfam; PF01842; ACT; 2.
 DR TIGRfam; TIGR00656; asp_kin_monofn; 1.
 DR TIGRfam; TIGR00857; asp_kinases; 1.
 DR PROSITE; PS00324; ASPARTOKINASE; 1.
 KW Transferrase; Kinase; Diaminopimelate biosynthesis;
 KW Lysine biosynthesis; Alternative initiation; Complete proteome.
 CC CHAIN 1 421 ASPARTOKINASE, ISOFORM ALPHA.

FT CHAIN 250 421 ASPARTOKINASE, ISOFORM BETA.
 FT INIT MET 250 250 FOR ISOFORM BETA.
 FT MUTAGEN 301 301 S->Y: FEEDBACK-RESISTANT AND ENHANCED
 FT EXPRESSION OF THE ASD GENE.
 FT CONFLICT 40 40 C -> V (IN REF. 1 AND 3).
 SQ SEQUENCE 421 AA; 44754 MW; E36B4D0081DE0827 CRC64;
 Query Match 55.8%; Score 1224.5; DB 1; Length 421;
 Best Local Similarity 64.0%; Pred. No. 8.le-79;
 Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
 QY 1 VQYGGSSVADAEIRIRVAAIRIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 DB VQYGGSSVADAEIRIRVAAIRIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 QY 5 VQYGGSSVADAEIRIRVAAIRIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPPPR 64
 DB VQYGGSSVADAEIRIRVAAIRIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPPPR 64
 QY 61 ELDMLTAGERISNALVMAIESLGARHSFTQSQAGVITTTGTHGNAKIIDVTGRLQTA 120
 DB ELDMLTAGERISNALVMAIESLGARHSFTQSQAGVITTTGTHGNARIVDVTPGRVRA 124
 QY 121 LFGRRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADYCEIYTDVIGIFSADP 180
 DB LFGRRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADYCEIYTDVIGIFSADP 180
 QY 125 LDEGKICIVAGFGVQNKETRDVTTLGRGSGDTTAVAMAAALGADYCEIYSDVDGVYADP 184
 DB LDEGKICIVAGFGVQNKETRDVTTLGRGSGDTTAVAMAAALGADYCEIYSDVDGVYADP 184
 QY 181 RIVNARKLDTVTFEEMLEMAACGKVLMLRCVEYARRHNI PVHVPSSYSDRPGTVVGS 240
 DB RIVNARKLDTVTFEEMLEMAACGKVLMLRCVEYARRHNI PVHVPSSYSDRPGTVVGS 240
 QY 185 RIVNARKLDTVTFEEMLEMAACGKVLMLRCVEYARRHNI PVHVPSSYSDRPGTVVGS 244
 DB RIVNARKLDTVTFEEMLEMAACGKVLMLRCVEYARRHNI PVHVPSSYSDRPGTVVGS 244
 QY 241 IKDVPMDPILTGVAHDSRSEAKVTIVGLPDIPOYAAKVFRAVA-----RRRQRH 290
 DB IKDVPMDPILTGVAHDSRSEAKVTIVGLPDIPOYAAKVFRAVA-----RRRQRH 290
 QY 245 MEDIPVEEAVLTGATDKSEAKVTIVGLPDIPOYAAKVFRAVA-----RRRQRH 304
 DB MEDIPVEEAVLTGATDKSEAKVTIVGLPDIPOYAAKVFRAVA-----RRRQRH 304
 QY 291 HGAERLQGRGROD-RHLLHLLP--QTSGPPKWKWTRSETRSASTOLLYDHIKVSLLI 347
 DB HGAERLQGRGROD-RHLLHLLP--QTSGPPKWKWTRSETRSASTOLLYDHIKVSLLI 347
 QY 305 DGTDTITFTCPSSDGRRAEILKQLVQG-----TNVLYDDQVGVKSLV 350
 DB DGTDTITFTCPSSDGRRAEILKQLVQG-----TNVLYDDQVGVKSLV 350
 QY 348 GAGMRSHPGVTATFCBALAAGVNIELISTSE 379
 DB GAGMRSHPGVTATFCBALAAGVNIELISTSE 379
 QY 351 GAGMRSHPGVTATFCBALAAGVNIELISTSE 382
 DB GAGMRSHPGVTATFCBALAAGVNIELISTSE 382
 RESULT 5
 AK CORFL STANDARD; PRT; 421 AA.
 ID AK CORFL
 AC P41398;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 13-SEP-2003 (Rel. 42, Last annotation update)
 DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
 DE alpha subunit; Aspartokinase beta subunit].
 GN LYSC OR ASK.
 OS Corynebacterium flammum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriidae; Corynebacterium.
 OX NCBI_TaxID=28028;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=N13;
 RX MEDLINE=93308089; PubMed=8100567;
 RA Follettie M.T., Peoples O.P., Agoropoulou C., Sinskey A.J.;
 RT "Gene structure and expression of the Corynebacterium flammum N13
 RT ask-asd operon.";
 RL J. Bacteriol. 175:4096-4103 (1993).
 CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
 CC aspartate.
 CC -!- ENZYME REGULATION: LOW FEEDBACK INHIBITION BY LYSINE AND
 CC THREONINE. THE DECREASED SENSITIVITY OF C. FLAMMUM, AS COMPARED TO
 CC C. GLUTAMICUM, IS PROBABLY DUE TO THE D345G SEQUENCE DIFFERENCE.
 CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
 CC FROM ASP TO THE CELL WALL. PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,
 CC TO MET, TO ILE AND TO THR.
 CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
 CC isoforms Beta (function not known) (By similarity).
 CC -!- ALTERNATIVE PRODUCTS.

```
CC Event=Alternative initiation;
CC Comment=2 isoforms, Alpha/Aspartokinase alpha subunit (shown
CC here) and Beta/Aspartokinase beta subunit, are produced by
CC alternative initiation;
CC -1- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L16848; AAA23293.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aakinas; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
DR TIGRFAMs; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
FT CHAIN 1 421
FT CHAIN 250 421
FT INIT_MET 250 250
FT MUTAGEN 345 345
FT ACTIVITY TO CONCERTED FEEDBACK INHIBITION
FT BY LYSINE AND THREONINE.
FT
SQ SEQUENCE 421 AA; 44796 MW; 4B37B0B6597F665E CRC64;
Query Match 55.5%; Score 1216.5; DB 1; Length 421;
Best Local Similarity 63.8%; Pred. No. 3e-78;
Matches 250; Conservative 55; Mismatches 60; Indels 27; Gaps 5;
QY 1 VQKYGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCVPPPPR 60
D 5 VQKYGSSLSAEIRIRVAERIVATKKAGNDVVVVVSAMGDTTDDLELAAVNPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMATESIGAHARSFTSGQAVITTTGTHGNAKIIDVTPEGLQTA 120
D 65 EMDMLLTAGERISNALVAMATESIGAEQAQFTSGQAVITTTGTHGNARIIDVTPEGLQTA 124
QY 121 LEEGRVVLVAGFQGVSDQTDVTTTLGRGGSDTTAVAMAAALGADVCEIYTDVDFGSADP 180
D 125 LDEGKICIVAGFQGVKNETRDVTTTLGRGGSDTTAVAAALNADVCIELYSDVDGVYTDAP 184
QY 181 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVYARRHNPVHVRSSYSDRGTVTVVGS 240
D 185 RIVNPAQKLEKLSPEEMLELAAGSKILVLRSEYARAFNPVLRVRSYSDNDPGLIAGS 244
QY 241 IKDVPMDPIITGVADHRSEAKVTIGVLPDIPGVAAKVFRAVA-----RRRQHR 290
D 245 MEDIPVEAVLTGVATDKSEAKVTIGLISDXPGEAAKVFRALADABINIDMVLQNVSSVE 304
QY 291 HGAERLQGRQD-RHHLHLLP--QTSPPPPKNNWTRSETRSASTOLLDDHIGKVSLLI 347
D 305 DGTDDITFTCPADGRRAAMEIKKLQVG-----NW-----TNVLYDDQDKVSLV 350
QY 348 GAGMRSHPGVTATCEALAAVGVNIELISTSE 379
D 351 GAGMRSHPGVTAEFMEALRDVNVNIELISTSE 382
RESULT 6
AK_PSEAE STANDARD; PRT; 412 AA.
AC O69077; Q91552;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
```

```
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase).
GN LYSC OR PA0904.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI TaxID=287;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437937; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman P.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [2]
RP SEQUENCE OF 198-412 FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RA Hindle Z., Throup J.P., Francis K.P., Bycroft B.W., Williams P.,
RA Stewart G.S.A.B.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -1- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE, TO MET, TO ILE AND TO THR.
CC -1- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF004525; AAG04293.1; -.
DR EMBL; AF061757; AAC16241.1; -.
DR PIR; C83531; C83531.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aakinas; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
DR TIGRFAMs; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Complete proteome.
SQ SEQUENCE 412 AA; 44388 MW; A06ACAA5E914F337 CRC64;
Query Match 37.5%; Score 823; DB 1; Length 412;
Best Local Similarity 42.6%; Pred. No. 1.2e-50;
Matches 170; Conservative 80; Mismatches 105; Indels 44; Gaps 4;
QY 1 VQKYGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCVPPPPR 60
D 5 VQKFGTSGVTVERIEQVAEKVKFREAGDDVVVVVSAMSGETNRLIGLANQIMEQVPR 64
QY 61 ELDMLLTAGERISNALVAMATESIGAHARSFTSGQAVITTTGTHGNAKIIDVTPEGLQTA 120
D 65 ELDVWVSTGEQVTIALLSNALIKRGVPFVSTGNQVRILTDSAKTKARLHIDDTTHRAD 124
QY 121 LEEGRVVLVAGFQGVSDQTDVTTTLGRGGSDTTAVAMAAALGADVCEIYTDVDFGSADP 180
D 125 LKAGRVVVVAGFQGV-DGNGNITTLGRGGSDITGVALAALKADECIYTDVGVYTTDP 183
QY 181 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVYARRHNPVHVRSSYSDRGTVTVVGS 240
```

```

Db 184 RVVPQARRLDKITFEEMLEMASIGSKVLQIRAVEFAGKYNVPLRVLHSGFOEGPGLTID 243
QY 241 IKDVPVEDPLITGVVADRSEAKVTIVGLPDIPGYAAKVFRAVARRRQ----- 288
Db 244 DEESMEQIISIAINRDEAKLTIRGVPDTPGVAFKILGPISAAANVEVDMIVQVADHN 303
QY 289 -----HRHGAERLOGRQDRHLLHLLPQTSGPPPKWQWTRSETRASTOLLYDDH 340
Db 304 TTDFTVHRNDYLALE-----ILAQTA-----ANIGAREALGDTN 340
QY 341 IGVSVLIGMRSHPGVTFATFCEALAAVGVNIELISTSE 379
Db 341 IAKVSVIGVMRSHAGVAGRNFEALAKESINQMISTSE 379

RESULT 7
AK_THETH STANDARD; PRT; 405 AA.
AC P97151; P77991;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
DE alpha subunit (ASK-alpha); Aspartokinase beta subunit (ASK-beta)].
GN ASK OR ASKAB.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB27;
RX MEDLINE=99173891; PubMed=10074061;
RA Kobashi N., Nishiyama M., Tanokura M.;
RA Nishiyama M., Kukimoto M., Beppu T., Horiuchi S.;
RT "An operon encoding aspartokinase and purine phosphoribosyltransferase
RT in thermophilic bacterium, Thermus thermophilus: lysine is synthesized
RT via alpha-amino adipic acid not via diaminopimelic acid.";
RL J. Bacteriol. 181:1713-1718 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33923 / DSM 674 / AT-62;
RX MEDLINE=95291465; PubMed=7773416;
RA Nishiyama M., Kukimoto M., Beppu T., Horiuchi S.;
RT "An operon encoding aspartokinase and purine phosphoribosyltransferase
RT in Thermus flavus";
RL Microbiology 141:1211-1219 (1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE, TO MET, TO ILE AND TO THR.
CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
CC isoforms Beta (function not known) (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative initiation;
CC Comment=2 isoforms, Alpha/Aspartokinase alpha subunit (shown
CC here) and Beta/Aspartokinase beta subunit, may be produced by
CC alternative initiation;
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB013131; BAA25848.1; -
CC EMBL; AB013131; BAA25849.1; -
CC EMBL; D37928; BAA07146.1; -
CC EMBL; D37928; BAA07147.1; -
CC InterPro; IPR001048; Aa_kinase.

```

```

DR InterPro; IPR002912; ACT.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aakinase; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRFAMS; TIGR00656; asp_kin_monofn; 1.
DR TIGRFAMS; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Alternative initiation.
FT CHAIN 1 405 ASPARTOKINASE, ISOFORM ALPHA.
FT CHAIN 245 405 ASPARTOKINASE, ISOFORM BETA.
FT INIT MET 245 245 FOR ISOFORM BETA.
FT VARIANT 126 126 E -> D (IN STRAIN AT-62).
SQ SEQUENCE 405 AA; 43333 MW; 06F7B79E908960EA CRC64;

Query Match 37.3%; Score 819; DB 1; Length 405;
Best Local Similarity 43.2%; Pred. NO. 2.2e-50;
Matches 175; Conservative 78; Mismatches 94; Indels 58; Gaps 6;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPPR 60
Db 5 VKYGGTSVGDLEIRIHKVAQIAHYEKEGRLAVVVSAMGHTTDLIALAKVNRPPFR 64
QY 61 ELDMLTAGERISNALVMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTGRLQTA 120
Db 65 ELDLTTTGEQSVALLSMQLWAMGIPAKGFVQHQTITTDGRYGDARILEVNPAREA 124
QY 121 LEGRVVLVAGFOGVSQDTKVTTLGRGSDTTAVAMAALGADVCEIYTDVDGIFSDAP 180
Db 125 LEQGFVAVIAGFMGTPE-GEITTLGRGSDTTAVAIALAKAKECEIYTDIEGVTTDP 183
QY 181 RIVNRARKLDTVTFEEMLENAACGKVMLCVVEYARRHNPVHVRSSYSRDPGTVWGS 240
Db 184 HLIPKARKLSVIGYDQMLEWALGARVLPRAVYAKRYGVVLHVRSSFSYNPGLTV--- 240
QY 241 IKDVPME-DPLITGVADRSEAKVTIVGLPDTPGYAAKVFRAVARR----- 285
Db 241 -KEVAMEMKAVTGVVALDLCHQIGLIGIPDOPGIAKVFQALAEAGIADVMIQGVPH 299
QY 286 -----RQHRHGAERLOGRQDRHLLHLLPQTSGPPPKWQWTRSETRASTQ 334
Db 300 DPSRQQAFTVKDPQAEALE-----PVLAIEGG-----E 332
QY 335 LLYDDHIGKVLIGACMRSHPGVTFATFCEALAAVGVNIELISTSE 379
Db 333 AILRPDIKXSVIVGVLASTPEVPAPKMFQAVASTGANIEMIATSE 377

RESULT 8
AK2_BACSP STANDARD; PRT; 411 AA.
AC Q99229;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase 2 (EC 2.7.2.4) (Aspartokinase II) (Aspartate kinase 2)
DE [Contains: Aspartokinase II alpha subunit; Aspartokinase II beta
DE subunit].
DE LysC.
GN Bacillus sp. (strain MGA3).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=93073880; PubMed=1444390;
RA Schendel F.J., Flickinger M.C.;
RT "Cloning and nucleotide sequence of the gene coding for aspartokinase
RT II from a thermophilic methylophilic Bacillus sp.";
RL Appl. Environ. Microbiol. 59:2806-2814 (1992).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- ENZYME REGULATION: LYSINE-SENSITIVE.

```

```
CC CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC CC FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
CC CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
CC CC isoforms Beta (function not known).
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event-Alternative initiation;
CC CC Comment=2 isoforms, Alpha/Aspartokinase II alpha subunit (shown
CC CC here) and Beta/Aspartokinase II beta subunit, are produced by
CC CC alternative initiation;
CC CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M93419; AAA22251.1; -.
CC CC InterPro; IPR001048; Aa_kinase.
CC CC InterPro; IPR002912; ACT.
CC CC InterPro; IPR005260; Asp_kin_monofn.
CC CC InterPro; IPR001341; Aspartate_kinase.
CC CC Pfam; PF00696; aakinas; 1.
CC CC Pfam; PF01842; ACT; 2.
CC CC TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
CC CC TIGRFAMs; TIGR00657; asp_kinases; 1.
CC CC PROSITE; PS00324; ASPARTOKINASE; 1.
CC CC Transferrase; Kinase; Diaminopimelate biosynthesis;
CC CC Lysine biosynthesis; Alternative initiation.
CC CC CHAIN 1 411 ASPARTOKINASE 2, ISOFORM ALPHA.
CC CC FT CHAIN 246 411 ASPARTOKINASE 2, ISOFORM BETA.
CC CC FT INIT MET 246 246 FOR ISOFORM BETA.
CC CC SQ SEQUENCE 411 AA; 44342 MW; 94061572E369CD48 CRC64;

Query Match 35.8%; Score 785; DB 1; Length 411;
Best Local Similarity 42.1%; Pred. No. 5.5e-48;
Matches 170; Conservative 81; Mismatches 97; Indels 56; Gaps 6;

Qy 1 VQKYGSSVADAEIRRVVAERIVATKKQGNVTVVVSAMGDTTDDLLDAQVCPAPPPR 60
Db 5 VQKFGTSGVSVIRILANVRVIEBKNGNDVWVVSAMGKTTDELVDLAKQISAPPKR 64

Qy 61 ELDMLITAGERISNALVAMATESLGAHARSFTGSCAGVITTGTHGNAKIIDVTPGLQTA 120
Db 65 EMDMLITGEGVITSLALWALNEKGYFAISTGQAGITTEPVFNARILNIETEKIQK 124

Qy 121 LEEGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAMAALGADVCEIYTDVDFGIFADP 180
Db 125 LNEGKIIVVAGFQGIIDEH-GEITTLGRGSGDTTAVALAALKAECIDIYDVTGVFTDP 183

Qy 181 RIVNRARLDTVTPEEMLEMAACAGKVLMLRCVEYARRHNPVHVRYSYSDRPGTVVGS 240
Db 184 RYVSARKLASISYDEMLELANLGVLPRAVFAKNYGYITLVRSSMEREEGTIIIE-- 241

Qy 241 IKQVPM-EPILTGVADHRSKATVIGLPIPGVAAKRVFVARRRRQRHGAERLQG 299
Db 242 -EVTMEQNLVVRGVAFDEITRVTFGLPNSLSLSTIFTTLA----- 284

Qy 300 RGRDRHLLHPQSGPPPPNNWTRSETRASQTLLYDD----- 339
Db 285 ---QNRINVDIIQSA-----TDAETTLNLSFIKSDDLLETMAVLENNKNLNYQGI 333

Qy 340 ---HIGKVSILGAGMRSPGVATFCEALAAVGNVIELISTSE 379
Db 334 BSETGLAKVISVGSNINPNPVAAKMFEVLALNGIQVKNVTSSE 377

RESULT 9
AK2_BACST
ID_AK2_BACST
AC_P53553;

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase 2 (EC 2.7.2.4) (Aspartokinase II) (Aspartate kinase 2)
DE [Contains: Aspartokinase II alpha subunit; Aspartokinase II beta
DE subunit].
GN LysC.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NB3621;
RX MEDLINE=96186919; PubMed=8635739;
RA Cantoni R., Labo M., de Rossi E., Riccardi G.;
RT "Sequence of the Bacillus stearothermophilus gene encoding
RT aspartokinase II.";
RL Gene 169:135-136(1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- ENZYME REGULATION: LYSINE-SENSITIVE (BY SIMILARITY).
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
CC isoforms Beta (function not known).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative initiation;
CC Comment=2 isoforms, Alpha/Aspartokinase II alpha subunit (shown
CC here) and Beta/Aspartokinase II beta subunit, may be produced by
CC alternative initiation;
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L46351; AAB06216.1; -.
CC FIR; JC4640; JC4640.
CC InterPro; IPR001048; Aa_kinase.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR005260; Asp_kin_monofn.
CC InterPro; IPR001341; Aspartate_kinase.
CC Pfam; PF00696; aakinas; 1.
CC Pfam; PF01842; ACT; 2.
CC TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
CC TIGRFAMs; TIGR00657; asp_kinases; 1.
CC PROSITE; PS00324; ASPARTOKINASE; 1.
CC Transferrase; Kinase; Diaminopimelate biosynthesis;
CC Lysine biosynthesis; Alternative initiation.
CC CHAIN 1 407 ASPARTOKINASE 2, ISOFORM ALPHA.
CC FT CHAIN 246 407 ASPARTOKINASE 2, ISOFORM BETA
CC (PROBABLE).
CC FT INIT MET 246 246 FOR ISOFORM BETA (PROBABLE).
CC SQ SEQUENCE 407 AA; 44002 MW; 2A848E6A7037B172 CRC64;

Query Match 35.5%; Score 778; DB 1; Length 407;
Best Local Similarity 43.2%; Pred. No. 1.7e-47;
Matches 173; Conservative 77; Mismatches 98; Indels 52; Gaps 7;

Qy 1 VQKYGSSVADAEIRRVVAERIVATKKQGNVTVVVSAMGDTTDDLLDAQVCPAPPPR 60
Db 5 VQKFGTSGVSVIRILANVRVIEBKNGNDVWVVSAMGKTTDELVDLAKQISAPPKR 64

Qy 61 ELDMLITAGERISNALVAMATESLGAHARSFTGSCAGVITTGTHGNAKIIDVTPGLQTA 120
Db 65 EMDMLITGEGVITSLALWALNEKGYFAVSLTGQAGITTEMHGNARINIDTTIRRC 124

Qy 121 LEEGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAMAALGADVCEIYTDVDFGIFADP 180
Db 125 LNEGKIIVVAGFQGIIDEH-GEITTLGRGSGDTTAVALAALKAECIDIYDVTGVFTDP 183
```

```

Db 125 LDEGAIVIVAGFGVGT-ETGEITTLGRGGSDTTAVALAALKAEXCDIYDVTGVTTPD 183
QY 181 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSDRPGTVVVG 240
Db 184 RYKTKARKIKETSYDEMELANLGAAGVLPRAVEFAKNYVEPLEVRSSMENERGTMRK- 241
QY 241 IKDVPMEDEP-ILITGVADRSEAKVTIVGLPDPGVAAKVFRVARRRRQHRHGAERLQG 299
Db 242 -EEVSEQHILVIRGIAFEQDVTRVTVGIEKVLQSVATIFTALANR----- 286
QY 300 RGRQDRHLLHLLPQTSPPPPKWNTRSETRSAS-----TOLL-----YDD 339
Db 287 -----GINVDIIIQ-----NATNSETASVSFSTRTEDLPETLQVLQALEGADVHVES 333
QY 340 HIGKVLIGAGMRSHRGVTATCEALAAVGVNIELISTSE 379
Db 334 GIAKSVISVSGMISNPGVAARFEVLADOGIEIKWVSISE 373

RESULT 10
AK_AQUAE STANDARD; PRT; 415 AA.
AC O67221;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase).
GN LYSC OR AQ 1152.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RA MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lennox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358 (1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE, TO MET, TO ILE AND TO THR.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000726; AAC07182.1; -
CC PIR; B70399; B70399.
CC InterPro; IPR001048; Aa kinase.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR005260; Asp_kin_monofn.
CC Pfam; PF00696; aakkinase; 1.
CC Pfam; PF01842; ACT; 2.
CC TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
CC TIGRFAMs; TIGR00657; asp_kinases; 1.
CC PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferrase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Complete proteome.
SQ SEQUENCE 415 AA; 45519 MW; E811750A8C14162 CRC64;

Query Match 33.5%; Score 734.5; DB 1; Length 415;
Best Local Similarity 42.4%; Pred. No. 1.9e-44;
Matches 164; Conservative 76; Mismatches 124; Indels 23; Gaps 6;

```

```

QY 1 VKYGGSSVADABRIRRVARRIVATKKQGNVTVVVSAMGDDTTDDLLDAQVCPPPP 60
Db 5 VQKFGTSTVGLTERIELCAKRVIKALEKGYKPYVVSAMAGETDSLIELAKKVHPPPP 64
QY 61 ELDMLTAGERINALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIDVTPGRLQTA 120
Db 65 EMTLLAVGEQKALAFMTLNKLGPAVSLCQWQVPIITDNVHTKARKIGISLLSL 124
QY 121 LEEGRVVLVAGFGVQSDTKQVTTLGRGGSDTTAVAMAALGADVCEIYTDVDFGSADP 180
Db 125 IEEGYPVWAGFGVQVTEDEW-EITTLGRGGSDTTAVALAALRAD-CEINTDVGFTADP 182
QY 181 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSDRPGTVVVG 240
Db 183 RIVNARKIKKISYEEMLEMAAGVMQIRSVFAAKYNVRIHVRSTFESEEGTWIVP- 241
QY 241 IKDVPMEDEPILITGVADRSEAKVTIVGLPDPGVAAKVFRVARRRRQHRHGA- 294
Db 242 -EEVMEKVAARGITVDTKTETRTVTVVPDKPGIAAEIIFKAL-----GDAHIVDM 291
QY 295 --RELQGRGDRHLLHLLPQTSPPPPKWNTRSETRSASTQLLYDDHIGKVLIGAGMR 352
Db 292 IVQNVSHGYTDM--FTVNTKDADKAEIVKIAQKIGAKEVVRDVKAKISVIGMK 349
QY 353 SHPGVTATCEALAAVGVNIELISTSE 379
Db 350 SAYGVAGRMFDILARNGINIKALISTSE 376

RESULT 11
AK_HELPU STANDARD; PRT; 405 AA.
AC Q9JZ7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase).
GN LYSC OR JHP1150.
OS Helicobacter pylori J99.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OX Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Urias-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merzberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180 (1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE, TO MET, TO ILE AND TO THR.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001542; AAD06723.1; -
CC PIR; H71843; H71843.
CC InterPro; IPR001048; Aa kinase.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR005260; Asp_kin_monofn.

```

```
DR InterPro: IPR001341; Aspartate_kinase.
DR Pfam: PF00696; aakinas; 1.
DR Pfam: PF01842; ACT; 2.
DR TIGRfam: TIGR00656; asp_kin_monofn; 1.
DR TIGRfam: TIGR00657; asp_kinases; 1.
DR PROSITE: PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Complete proteome.
SQ SEQUENCE 405 AA; 44219 MW; E43A0D62E890A8EF CRC64;

Query Match 32.9%; Score 721; DB 1; Length 405;
Best Local Similarity 41.4%; Pred. NO. 1.7e-43;
Matches 161; Conservative 84; Mismatches 120; Indels 24; Gaps 6;

QY 1 VKYGGSSVADAEIRIRVEARIVATKKQGNVVVVVSAMGDTTDDLLDLAQVCVAPP 60
DB 1 VKYGGSSVADAEIRIRVEARIVATKKQGNVVVVVSAMGDTTDDLLDLAQVCVAPP 60
QY 4 VKYGGTSGMSIERIHVQAQVLESVKLGQVQVVVSAMSGETDLRLEFGKFNPNKR 63
DB 4 VKYGGTSGMSIERIHVQAQVLESVKLGQVQVVVSAMSGETDLRLEFGKFNPNKR 63
QY 61 ELMILTAGERISNALVAVAIISLGHARSFTGCSQAGVTTGTHGNAKIIVTPQLQTA 120
DB 61 ELMILTAGERISNALVAVAIISLGHARSFTGCSQAGVTTGTHGNAKIIVTPQLQTA 120
QY 64 EMDRIVSAGEWISSAALSVALERYGHRALSLSGKEAGILTSFHFQNAVQISDTQITEL 123
DB 64 EMDRIVSAGEWISSAALSVALERYGHRALSLSGKEAGILTSFHFQNAVQISDTQITEL 123
QY 121 LBEGRVVLVAGFGVGSQDITKDTTLGRGSDTTAVAMAALGADVCEIYTDVDFPSADP 180
DB 121 LBEGRVVLVAGFGVGSQDITKDTTLGRGSDTTAVAMAALGADVCEIYTDVDFPSADP 180
QY 124 LEKNYIVVIAGFQG-ADIQETTTLGRGSDLSVALAGALKALHCEIYTDVGVYITDP 182
DB 124 LEKNYIVVIAGFQG-ADIQETTTLGRGSDLSVALAGALKALHCEIYTDVGVYITDP 182
QY 181 RIVRNARKLDTVTFFEMLEMAACCAKVLMLRCVEYARHNPVHVSSVSDRPGTVVVG- 239
DB 181 RIVRNARKLDTVTFFEMLEMAACCAKVLMLRCVEYARHNPVHVSSVSDRPGTVVVG- 239
QY 183 RIBEKAQKIAQISYDEMLELASKGAKVLLNRSVELAKLSVKLVTSNFSHSEGLIVAE 242
DB 183 RIBEKAQKIAQISYDEMLELASKGAKVLLNRSVELAKLSVKLVTSNFSHSEGLIVAE 242
QY 240 -STKVDPMEDPILTVGAHDBSEAKVTVGLPDIPGVAAKVRAVARRRQRHGAERLQ 298
DB 240 -STKVDPMEDPILTVGAHDBSEAKVTVGLPDIPGVAAKVRAVARRRQRHGAERLQ 298
QY 243 KDFKGERMETPTVSGIALDKQARVMEGVDEDPGIAAEIFGALA-----EYRINDVMIVQ 298
DB 243 KDFKGERMETPTVSGIALDKQARVMEGVDEDPGIAAEIFGALA-----EYRINDVMIVQ 298
QY 299 GRGQRDRHHL-LPQTSQPPPPKQWTRSTRSTRASTQLL-----YDDHICKVSLIGAG 350
DB 299 GRGQRDRHHL-LPQTSQPPPPKQWTRSTRSTRASTQLL-----YDDHICKVSLIGAG 350
QY 299 TIGRDKGKTDLDFTIVKQ-----IETKQALKPFLAQMDSIDYDENIAKVISVGVG 349
DB 299 TIGRDKGKTDLDFTIVKQ-----IETKQALKPFLAQMDSIDYDENIAKVISVGVG 349
QY 351 MRHPGVWTFATFCALAAVGNIELISTSE 379
DB 351 MRHPGVWTFATFCALAAVGNIELISTSE 379
QY 350 MKSHSGVASIAFKALAKDNINIMISTSE 378
DB 350 MKSHSGVASIAFKALAKDNINIMISTSE 378

RESULT 12
AK2_BACSU STANDARD; PRT; 408 AA.
ID -AK2_BACSU STANDARD; PRT; 408 AA.
AC P08495; P08496; P94554; P97183;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase 2 (EC 2.7.2.4) (Aspartokinase II) (Aspartate kinase 2)
DE [Contains: Aspartokinase II alpha subunit; Aspartokinase II beta subunit].
GN LVSC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87250503; PubMed=3036830;
RA Chen N.-Y., Hu F.M., Paulus H.;
RA "Nucleotide sequence of the overlapping genes for the subunits of Bacillus subtilis aspartokinase II and their control regions.";
RL J. Biol. Chem. 262:8787-8798(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=97124191; PubMed=8969504;
RA Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.,
RA Sanders J., Emmerson P.T., Harwood C.R.;
RA "The dnaB-phaA (256 degrees-240 degrees) region of the Bacillus subtilis chromosome containing genes responsible for stress responses, the utilization of plant cell walls and primary metabolism.";
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azavedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Enrian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Jovis B., Katamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha B., Roche B., Rose M., Sadale Y.,
RA Sato T., Scallan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP ALTERNATIVE INITIATION.
RX MEDLINE=88243847; PubMed=2837491;
RA Chen N.-Y., Paulus H.;
RT "Mechanism of expression of the overlapping genes of Bacillus subtilis aspartokinase II.";
RL J. Biol. Chem. 263:9526-9532(1988).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-aspartate.
CC -!- ENZYME REGULATION: LYSINE-SENSITIVE.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2 isoforms Beta (function not known).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative initiation;
CC Comment=2 isoforms, Alpha/Aspartokinase II alpha subunit (shown here) and Beta/Aspartokinase II beta subunit, are produced by alternative initiation;
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J03294; AAA87318.1; -
CC EMBL; J03294; AAA87319.1; -
CC EMBL; Z75208; CAA99580.1; -
CC EMBL; Z75208; CAA99581.1; -
CC EMBL; Z99118; CAB14807.1; -
CC PIR; A29314; A29314.
CC Subtilist; BG10350; lysc.
CC InterPro; IPR01048; Aa_kinase.
```


30-MAY-2000 (Rel. 39, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartokinase 1 (EC 2.7.2.4) (Aspartokinase 1) (Aspartate kinase 1)
[Contains: Aspartokinase I alpha subunit; Aspartokinase I beta subunit].
DE DAPG OR LSSP.
GN Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=1421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93252813; PubMed=8098035;
RA Chen N.-Y., Jiang S.-Q., Klein D.A., Paulus H.,
RT "Organization and nucleotide sequence of the Bacillus subtilis
RT diaminopimelate operon, a cluster of genes encoding the first three
RT enzymes of diaminopimelate synthesis and diaminopimelate synthase.",
RL J. Biol. Chem. 268:9448-9465(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
RA Arevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boursier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschini C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guilseppi G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kaeahara Y., Klaert-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kunano M.,
RA Kurita K., Lapicque A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rev M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadade Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Teateo V., Uchiyama S., Vanderbol M., Vannier P., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenssger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
CC -1- CATALYTIC ACTIVITY: AMP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -1- ENZYME REGULATION: DIAMINOPIMELATE-SENSITIVE.
CC -1- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
CC -1- SUBUNIT: Tetramer consisting of two alpha (catalytic) and two beta
CC (FUNCTION NOT KNOWN) chains (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L08471; AAA22384.1; -
CC EMBL; Z99112; CAB13549.1; -
CC PIR; C46665; C46665.

DR Subtilist; BG10784; dapG.
DR InterPro; IPR001048; As_kinase.
DR InterPro; IPR00260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aakinas; 1.
DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
DR TIGRFAMs; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Complete proteome.
FT CHAIN 1 404 ASPARTOKINASE I BETA SUBUNIT
FT CHAIN 245 404 ASPARTOKINASE I BETA SUBUNIT
FT CHAIN 399 399 E -> V (IN REF. 1).
FT CONFLICT 399 399 E -> V (IN REF. 1).
SQ SEQUENCE 404 AA; 42978 MW; 49A6DA70D70047C4 CRC64;
Query Match 24.3%; Score 532; DB 1; Length 404;
Best Local Similarity 36.5%; Pred. No. 3e-30;
Matches 145; Conservative 62; Mismatches 122; Indels 68; Gaps 11;
QY 1 VQYGGSSVADAERIRRAERIVATKQGNVVDVVVVSAMGD-----TTDDLDDLAQQVCP 55
DB 5 VQKGGTSVKDDKGRKALGHIKEAISEGYKVVVVVSNMGRKGPYATDSLGLLYGQDS 64
QY 56 APPRELDMLTAGERISNALVAIBSLGAHARSFTGSQAGVITTGTHGNAKIIDVTPG 115
DB 65 AISPREQDILLSGGETISSVFTSMLLDNGVKAALITCAQAQGLTNDQHTNAKIEMKPE 124
QY 116 RLQTALEGRVYVAGFGVQSDIKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDDGI 175
DB 125 RLFSVLANDHADVAVAGFQCAT-EGKDTTIGRGSDTSAALGAADVADYDITFDVEGV 183
QY 176 PSADPRIVRNARKLDTVFEEMLNACAKVLMRCVEYARRHNPVHVRSYSDRPGT 235
DB 184 MTADPRVVENAKPLPVVITYTEICNLAYQGAQVSPRAVEIAQAKVPVRVSTYSNDKGT 243
QY 236 VV-----VGSIKDVPNEDILITGAH--DSEAKVTI-VGLDPIQVAAKVFVAVRR 285
DB 244 LVTSHSSKVGVS-----DVFERLITGIAHKVDQTKVPAKIQYNNV---QTEVFKAMA-- 294
QY 286 RQHRHGAERLQGRGRQRHLLPQTSGPPPMKWNRSE-----TRSASTQLL 336
DB 295 -----NAGISVDFFNITPSEIYVTVAGNKTETAQRILM 327
QY 337 ---YDHI-----GKVSLLIGAGMRSHPGVATFCEALA 366
DB 328 DMGYDPMVTRNCALKVSAVGAGIMGVPGVTSKIVSALS 364
RESULT 15
AK_METUA STANDARD; PRT; 473 AA.
ID AK_METUA
AC Q57891;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable aspartokinase (EC 2.7.2.4) (Aspartate kinase).
GN M00571.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2561 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.D., Geoghegan N.S.M., Weidman J.E., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 20.8635 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-4
Perfect score: 2193
Sequence: 1 VQYGGSSVADAEIRRAE.....SAATRPCTRGDGRWACQ 424

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258032604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1545.5	70.5	421	16 Q9CB77	Q9cb77 mycobacteri
2	1272.5	58.0	421	2 Q8KQ29	Q8kq29 streptomyc
3	1267.5	57.8	421	2 Q9RQ25	Q9rq25 amycolatops
4	1232.5	56.2	421	2 Q93C57	Q93c57 corynebacte
5	1228.5	56.0	421	2 Q9AE85	Q9ae85 nocardia la
6	1224.5	55.8	421	2 Q93C55	Q93c55 corynebacte
7	1141	52.0	425	16 Q9XAI7	Q9xai7 streptomyc
8	1127	51.4	423	2 Q8GIQ5	Q8giq5 streptomyc
9	821	37.4	434	2 Q93TX4	Q93tx4 stigmatella
10	816.5	37.2	600	16 P74569	P74569 synchocyst
11	803	36.6	405	15 Q8F865	Q8f865 leptospira
12	798.5	36.4	408	2 Q8RU02	Q8ru02 methylobaci
13	786.5	35.9	599	16 Q8DHW0	Q8dhw0 synchococc
14	776.5	35.4	606	16 Q8YR14	Q8yr14 anabaena sp
15	773.5	35.3	254	16 Q8G6V6	Q8g6v6 bifidobacte
16	761	34.7	405	16 Q9JTN3	Q9jtn3 neisseria m

17	760	34.7	405	15 Q9JYN6	Q9jyn6 neisseria m
18	757.5	34.5	412	16 Q9K8B0	Q9k8b0 bacillus ha
19	742.5	33.9	400	16 Q9PHI4	Q9ph4 campylobact
20	724.5	33.0	401	16 Q9XIK6	Q9xik6 thermotoga
21	724.5	33.0	739	16 Q9WZ17	Q9wz17 thermotoga
22	707.5	32.3	416	16 Q8Y073	Q8y073 ralstonia s
23	707	32.2	404	16 Q92CH4	Q92ch4 listeria in
24	705	32.1	415	16 Q9A9W8	Q9a9w8 caulobacter
25	703	32.1	404	16 Q8Y7N9	Q8y7n9 listeria mo
26	694.5	31.7	423	16 Q8Y097	Q8y097 brucella me
27	690.5	31.5	423	16 Q8FYK1	Q8fyk1 brucella su
28	664.5	30.3	418	2 Q8EV19	Q8ev19 bradyrhizob
29	663	30.2	417	16 Q88G91	Q88g91 rhizobium l
30	656	29.9	431	16 Q8U8C0	Q8u8c0 agrobacteri
31	652.5	29.8	424	16 Q92MK3	Q92mk3 rhizobium m
32	632	28.8	395	16 Q9KU44	Q9ku44 vibrio chol
33	605	27.6	418	16 Q8E8S2	Q8e8s2 shewanella
34	602	27.5	398	16 Q8RA48	Q8ra48 thermococ
35	585	26.7	370	16 Q8DC48	Q8dc48 vibrio vuln
36	585	26.7	401	16 Q99U91	Q99u91 staphylococ
37	584	26.6	401	2 Q9EV36	Q9ev36 staphylococ
38	584	26.6	401	16 Q8NW57	Q8nws7 staphylococ
39	575	26.2	410	16 Q8CSM8	Q8csm8 staphylococ
40	562	25.6	411	16 Q9KA90	Q9ka90 bacillus ha
41	536	24.4	411	16 Q8EQT0	Q8eqt0 oceanobacil
42	524	23.9	403	16 Q92BR9	Q92br9 listeria in
43	520	23.7	401	16 Q92GF9	Q92gf9 rickettsia
44	506	23.1	403	16 Q8Y765	Q8y765 listeria mo
45	467	21.3	398	16 Q8XJ56	Q8xj56 clostridium

ALIGNMENTS

RESULT 1

Q9CB77	Q9CB77	PRELIMINARY;	PRT;	421 AA.
ID	Q9CB77			
AC	Q9CB77			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Aspartokinase.			
GN	ASK OR ML2323.			
OS	Mycobacterium leprae.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1769;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=IN;			
RX	MEDLINE=21128732; PubMed=11234002;			
RA	Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,			
RA	Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,			
RA	Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,			
RA	Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,			
RA	Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,			
RA	Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,			
RA	Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,			
RA	Barrell B.G.;			
RA	Nature 409:1007-1011(2001).			
RL	Nature 409:1007-1011(2001).			
DR	EMBL; AL583325; CAC31839.1;			
DR	Leprona; ML2323;			
DR	InterPro; IPR001048; Aa_kinase.			
DR	InterPro; IPR002912; ACT.			
DR	InterPro; IPR001341; Aspartate_kinase.			
DR	Pfam; PF00696; aakinas;			
DR	Pfam; PF01842; ACT.			
DR	TIGRams; TIGR00657; asp_kinases; 1.			
DR	PROSITE; PS00324; ASPARTOKINASE; 1.			
KW	Kinase; Complete proteome.			
SQ	SEQUENCE 421 AA; 44663 MW; 4A3BA44D1CEEPE06 CRC64;			

Query Match 70.5%; Score 1545.5; DB 16; Length 421;
 Best Local Similarity 80.4%; Pred. No. 3e-96;
 Matches 319; Conservative 18; Mismatches 43; Indels 17; Gaps 3;

QY 1 VOKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQOVCPPPPR 60
 DB 5 VOKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQOVCPPPPR 64
 QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPGRLOTA 120
 DB 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPGRLOTA 124
 QY 121 LEEGRVVLVAGQGVSDTKVTTTLGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 180
 DB 125 LDEGRVVLVAGQGVSDTKVTTTLGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 184
 QY 181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 240
 DB 185 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 244
 QY 241 IKDVPMEDELITGVADHRSSEAKVTIVGPDIPGYAAKVFRAVARRRRQHRHGAEE----- 295
 DB 245 IKDVPMEDELITGVADHRSSEAKVTIVGPDIPGYAAKVFRAVARRRRQHRHGAEE-----DADVNIDM 295
 QY 296 RLQGRGRDRHLLHLLPOTSGPPPKWNTSRSETRASTOLLYDDHIGKVSILGAMR 352
 DB 296 VLQNVSKVEDGKTDTFTCSRDGSLIAVAKGSLRDEIGFTOLLYDDHIGKVSILGAMR 355
 QY 353 SHPGVATFCEALAAVGNIELISTSEQRGCCAAT 389
 DB 356 SHPGVATFCEALAAVGNIELISTSEIRISVLCRDT 392

RESULT 2

Q8KQ29 ID Q8KQ29 PRELIMINARY; PRT; 421 AA.
 AC Q8KQ29;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Aspartate kinase alpha subunit (EC 2.7.2.4) (Aspartokinase).
 GN ASKA.
 OS Streptomyces clavuligerus.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycetaceae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1901;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tunica S.;
 RT "Isolation and characterization of the aspartokinase and aspartate semialdehyde dehydrogenase genes from cephamycin C-producer Streptomyces clavuligerus";
 RL Thesis (2002), University of Middle East Technical University, Ankara, Turkey.
 CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-ASPARTATE.
 CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
 CC EMBL: AY112728; AAM54736.1; -;
 DR InterPro; IPR001048; Aa_kinase.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR001341; Aspartate_kinase.
 DR InterPro; IPR005260; Asp_kin_monofn.
 DR InterPro; IPR001057; Glu_kinase.
 DR Pfam; PF00696; aak_kinase; 1.
 DR Pfam; PF01842; ACT; 2.
 DR PRINTS; PR00474; GLUSKINASE.
 DR TIGRFSMS; TIGR00657; asp_kinases; 1.
 DR TIGRFSMS; TIGR00656; asp_kin_monofn; 1.
 DR PROSITE; PS00324; ASPARTOKINASE; 1.
 KW Kinase; Transferase.
 SQ SEQUENCE 421 AA; 44399 MW; 08262D81045735C1 CRC64;

Query Match 58.0%; Score 1272.5; DB 2; Length 421;
 Best Local Similarity 66.1%; Pred. No. 7.5e-78;
 Matches 257; Conservative 51; Mismatches 60; Indels 21; Gaps 4;

QY 1 VOKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQOVCPPPPR 60
 DB 5 VOKYGGSSLESADRIKRAERIVATKAGNEVVVVVCSAMGDTTDELLDLAQOVPVPPPR 64
 QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPGRLOTA 120
 DB 65 EMDMLLTAGERISNALVAMAIESAQGAQNSFTGSGAGVTTTAVHGNARIIDVTPSEVSA 124
 QY 121 LEEGRVVLVAGQGVSDTKVTTTLGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 180
 DB 125 LEOGYVALVAGQGVADTKNDITTLGRGSDTTAVALAAALNADVCEIYSDVDGYVSADP 184
 QY 181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 240
 DB 185 RIVPNARKLDSVPEEMLELAASGSKILHRSVEYARRYGVFIRVRSSYSDRPGTVVVG 244
 QY 241 IKDVPMEDELITGVADHRSSEAKVTIVGPDIPGYAAKVFRAVARRRRQHRHGAEE----- 295
 DB 245 IEEIPVEQALITGVADHRSSEAKVTIVGPDHAGAAARIPRVIA-----DAEIDIM 295
 QY 296 RLQ-----GRGRDRHLLHLLPOTSGPPPKWNTSRSETRASTOLLYDDHIGKVSILGAG 350
 DB 296 VLQNVNTSGRTD--ITTLKANGPKAVASLEKIEELGFSVLYDDHVGKVSILVAGG 353
 QY 351 MRSHPGVATFCEALAAVGNIELISTSE 379
 DB 354 MRSHPGVATFCEALSKVGNIELISTSE 382

RESULT 3

Q9RQ25 ID Q9RQ25 PRELIMINARY; PRT; 421 AA.
 AC Q9RQ25;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Aspartokinase subunit A (EC 2.7.2.4) (Aspartate kinase).
 GN ASKA.
 OS Amycolatopsis mediterranei (Nocardia mediterranei).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
 OX NCBI_TaxID=33910;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=U32;
 RX MEDLINE=99453302; PubMed=10521665;
 RA Zhang W., Jiang W., Zhao G., Yang Y., Chiao J.;
 RT "Sequence analysis and expression of the aspartokinase and aspartate semialdehyde dehydrogenase operon from rifamycin SV-producing amycolatopsis mediterranei";
 RL Gene 237:413-419 (1999).
 CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-ASPARTATE.
 CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
 CC EMBL: AF134837; RAD49567.1; -;
 DR InterPro; IPR001048; Aa_kinase.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR001341; Aspartate_kinase.
 DR InterPro; IPR005260; Asp_kin_monofn.
 DR InterPro; IPR001057; Glu_kinase.
 DR Pfam; PF00696; aak_kinase; 1.
 DR Pfam; PF01842; ACT; 2.
 DR PRINTS; PR00474; GLUSKINASE.
 DR TIGRFSMS; TIGR00657; asp_kinases; 1.
 DR TIGRFSMS; TIGR00656; asp_kin_monofn; 1.
 DR PROSITE; PS00324; ASPARTOKINASE; 1.
 KW Kinase; Transferase.
 SQ SEQUENCE 421 AA; 44393 MW; 633D9C2D023145E9 CRC64;

Query Match		57.8%; Score 1267.5; DB 2; Length 421;
Best Local Similarity		65.3%; Pred. No. 1.6e-77;
Matches 254; Conservative 53; Mismatches 61; Indels 21; Gaps 4;		
QY	1	VQYGGSSVADAERIRRVAAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB	5	VQYGGSSLESABERIRRVAAERIVATKKAGNDVVVVCSAMGDTTDELLELAANVPVPPAR 64
QY	61	ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIVDTPGRLOTA 120
DB	65	EMDMLLTAGERISNALVAMAIESLGAESAQSFSGAGVLTTRHGNARIVDTPGRVREA 124
QY	121	LEEGRVVLVAGFGVGSODTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVGDGIFPSADP 180
DB	125	LDEGKICIVAGFGVGNKRETRDVTTLGRGSDTTAVALLAALNADVCEIYSDVGVYTADP 184
QY	181	RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSRDPGTVVVGS 240
DB	185	RIVPNAQKLEKLSFEEMLEAALVGSKILVLSRVSVEYARAFNVPLVRSSYSNDPGLIAGS 244
QY	241	IKDVPMDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRROHR 290
DB	245	MEDIPVEEAVLTGVTADKSEAKVTIVGLSDKPGGAACVFRALADAEINIDVQLNVSSVE 304
QY	291	HGAERLQCRQD--RHHLL--POTSGPPPKNWTSETRSASTOLLYDHGKVSLLI 347
DB	305	DGTTDTITCTPRSDGRRAMEILKKPQVG-----NW-----TNVLYDDQVGVSLV 350
QY	348	GAGMESHGVTATFCALAAVGVNIELISTSE 379
DB	351	CAGMKSHGVTAEFMEALRVNVIELISTSE 382
RESULT 5		
Q9AE85		PRELIMINARY; PRT; 421 AA.
ID	Q9AE85	
AC	Q9AE85	
DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	Aspartokinase alpha subunit (EC 2.7.2.4) (Aspartate kinase).	
GN	ASK.	
OS	Nocardia lactamdurans.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.	
OX	NCBI_TaxID=1913;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Hernando-Rico V.	
RL	Thesis (2000), Department of Area of Microbiology, University of Leon,	
RL	Leon, Spain.	
CC	-!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-ASPARTATE.	
CC	-!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.	
DR	EMBL; A0298904; CAC37034.1; -.	
DR	InterPro; IPR001048; Aa_kinase.	
DR	InterPro; IPR002912; ACT.	
DR	InterPro; IPR001341; Aspartate kinase.	
DR	InterPro; IPR001057; Glu_5kinase.	
DR	Pfam; PF00696; aakkinase; 1.	
DR	Pfam; PF01842; ACT; 2.	
DR	PRINTS; PRO00474; GLUSKINASE.	
DR	TIGRFAMS; TIGR00657; asp_kinases; 1.	
DR	PROSITE; PS00324; ASPARTOKINASE; 1.	
KW	Kinase; Transferase.	
SQ	SEQUENCE 421 AA; 44141 MW; 802FBE02D01B7F95 CRC64;	
Query Match		56.0%; Score 1228.5; DB 2; Length 421;
Best Local Similarity		64.3%; Pred. No. 7e-75;
Matches 250; Conservative 52; Mismatches 66; Indels 21; Gaps 4;		
QY	1	VQYGGSSVADAERIRRVAAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB	5	VQYGGSSLESADRIKRVAAERIVATKKAGNEVVVVCSAMGETTDELLELAQVNPVPGA 64
QY	61	ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIVDTPGRLOTA 120
Query Match		56.2%; Score 1232.5; DB 2; Length 421;
Best Local Similarity		64.3%; Pred. No. 3.7e-75;
Matches 252; Conservative 55; Mismatches 58; Indels 27; Gaps 5;		
QY	1	VQYGGSSVADAERIRRVAAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB	5	VQYGGSSLESADRIKRVAAERIVATKKAGNEVVVVCSAMGETTDELLELAQVNPVPGA 64
QY	61	ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIVDTPGRLOTA 120

```

Db 65 EMDMLTAGDGSINALVAVAIQAQAQWSTFGSQAGVVVTTAVHGNARIIDVTPSRVSR 124
QY 121 LBEGRVLVAGFGQVSQDTKDVTTLGRGSGDITAVAMAAALGADYCEIYTDVDFGSADP 180
Db 125 LQGYVALVAGFGQVADTKDITTLGRGSGDITAVAMAAALNADYCEIYSDVDGYSADP 184
QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIIPVHVRSSYSDRPGTVVGS 240
Db 185 RIVDPARKLDSVPYEEMLEMAAGSKILHLRSVESARRYGVPIRVSSYSDKPGTITVGS 244
QY 241 IKDVPMEDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRRQHR 290
Db 245 IBEIPVEQALITGVADHRSEAKVTIVGPDHAGAAIRFRVIA-----DAEIDIDM 295
QY 296 RLQ-----GRGQRDRHLLHLLPQTSGPPPKWNTSRSTRSASTOLLYDDHIGKVSILGAG 350
Db 296 VLQNVSTSSGRD--ITFLSKANGPKRAVASLEKIKELGFSVLYDDHVGKVSILVAG 353
QY 351 MRSHPGVATFCEALAAVGNIELISTSE 379
Db 354 MRSHPGVATFCEALCKGVNIEINTSE 382

RESULT 6
Q93C55 PRELIMINARY; PRT; 421 AA.
AC Q93C55;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Aspartokinase LysC alpha subunit (EC 2.7.2.4) (Aspartate kinase).
OS Corynebacterium crenatum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=169810;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1542;
RA Liu Y., Ding J., Wang J.;
RT "Cloning and sequence analysis of aspartokinase genes from Corynebacterium crenatum."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-ASPARTATE.
CC -! SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
DR EMBL; AF414085; AAL07809.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate_kinase.
DR InterPro; IPR005260; Asp_kin_monofn.
DR Pfam; PF00696; aakinas; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRfams; TIGR00657; asp_kinases; 1.
DR TIGRfams; TIGR00656; asp_kin_monofn; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ _SEQUENCE 421 AA; 44754 MW; E36B4D008IDE0627 CRC64;

Query Match 55.8%; Score 1224.5; DB 2; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.3e-74;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 1 VQYGGSSVADAERIRRVARRIVATKKQNDVVVVSAMGDTTDDLDAQOVCPAPP 60
Db 5 VQYGGSSLESARIRNVAERIVATKKAGNDVVVVSAMGDTTDELEAAVNPVPPAR 64
QY 61 ELDMLLTAGERISNALVAVAIQAQAQWSTFGSQAGVVVTTAVHGNARIIDVTPSRVSR 120
Db 65 EMDMLLTAGERISNALVAVAIQAQAQWSTFGSQAGVVVTTAVHGNARIIDVTPSRVSR 124
QY 121 LBEGRVLVAGFGQVSQDTKDVTTLGRGSGDITAVAMAAALGADYCEIYTDVDFGSADP 180

```

```

Db 125 LBEGRVLVAGFGQVSQDTKDVTTLGRGSGDITAVAMAAALNADYCEIYSDVDGVTADP 184
QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIIPVHVRSSYSDRPGTVVGS 240
Db 185 RIVDPARKLDSVPYEEMLEMAAGSKILHLRSVESARRYGVPIRVSSYSDKPGTITVGS 244
QY 241 IKDVPMEDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRRQHR 290
Db 245 IBEIPVEQALITGVADHRSEAKVTIVGPDHAGAAIRFRVIA-----DAEIDIDM 295
QY 296 RLQ-----GRGQRDRHLLHLLPQTSGPPPKWNTSRSTRSASTOLLYDDHIGKVSILGAG 350
Db 296 VLQNVSTSSGRD--ITFLSKANGPKRAVASLEKIKELGFSVLYDDHVGKVSILVAG 353
QY 351 MRSHPGVATFCEALAAVGNIELISTSE 379
Db 354 MRSHPGVATFCEALCKGVNIEINTSE 382

RESULT 7
Q9XAI7 PRELIMINARY; PRT; 425 AA.
AC Q9XAI7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative aspartokinase.
OS ASK OR SCO3615 OR SC66T3.26C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomyces; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RC MEDLINE=97000351; PubMed=8943436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RC MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S., Rabinowitz B., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J., Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL539117; CAB45482.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate_kinase.
DR InterPro; IPR005260; Asp_kin_monofn.
DR Pfam; PF00696; aakinas; 1.

```



```

Db      81 EQDVVATGEQVSGIGVALAIQTKRKAVSLGHQVRIITDSTFSKARIKSIDAERIVDA 140
QY      121 LEEGRVVLVAGQGVSDTKQVTTTGRGSDTTAVMAAALGADVCEIYTDVDFGSADP 180
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      141 LKKQIVVAVGQGG-QDEHGNVTTTGRGSDTTAVMAAALGADACACETIYTDVGVYTTDP 199
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 240
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      200 NMVPAARKLERIAYEMLELASGVAKVQIRSVFAMKYKPLWVKSFTDDPGTILVCE- 258
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      241 IKDVPWEDPILTVGAHDSKAKVTIVGLPDIPGYAAKVFRAVAR- - - - -RQHRHG 292
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      259 -EDAAMEDVLVSGIAYDKNEAKIARIGVDFVGVAAKIFGALDEQSIIVVDLIVQVNSKDG 317
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      293 AAERLQGRQDRHLLHLLPQTSQPPPKWNTSRSTRASTQ-LLYDDHIGKVLIGAGM 351
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      318 RTDVTFTVGKAD- - - - -LTKA- - - - -KDVVKIATKSVKAEVETDDQVSKSVIGVGM 365
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      352 RSHPGVTATFCEALAAVGNIELISTSDQSRSCCAATPNWT 393
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      366 RNHSGVAAKMTVLAGAGVWQVMTSEIKVS- - - - -CVIHSNYT 405
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 10

```

P74569 PRELIMINARY; PRT; 600 AA.
AC Q8F865;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Aspartate kinase.
GN LYSC OR SUF0657.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asanizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL; D90916; BAA18676.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate_kinase.
DR InterPro; IPR005260; Asp_kin_monofn.
DR Pfam; PF00696; aak_kinase; 1.
DR Pfam; PF01842; ACT; 4.
DR TIGRFAMs; TIGR00657; asp_kinases; 1.
DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Complete proteome.
SQ SEQUENCE 600 AA; 63533 MW; B30B7101E9FC4D46 CRC64;

```

```

Query Match 37.2%; Score 816.5; DB 16; Length 600;
Best Local Similarity 43.9%; Pred. No. 6.6e-47;
Matches 183; Conservative 83; Mismatches 118; Indels 33; Gaps 7;

QY      1 VQYGGSSVADAEIRIRVAERIVATKQGNVTVVVSAMGDTTDDLDAQOVCPPPPR 60
Db      5 VQKGGTSVGTVERIQAVKRIKTVQGNLVVVVSAMGKSTDLVLAQKQISPNPCR 64
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSAGVITGTHGNNAKIIDVTPGRLOTA 120
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      65 EMDMLLTGEGQVSLALLSLAQEIDQPAISLTGAVGIVTFAEHSRARIIEIPDRLEHH 124
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      121 LEEGRVVLVAGQGVSDTKQVTTTGRGSDTTAVMAAALGADVCEIYTDVDFGSADP 179
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Db      125 LREKVVVAVGQGVSDTSSVEHLEITTLGRGSDTSAVALAAALKADFCETIYDVPGLITTD 184
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      180 PRIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 239
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      185 PRIVPEAQLMAEITCDEMELASLGAKVLHPRAVEIARNYGIPLVVRSSWSDEPGTKVA 244
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      240 SIKDVPWEDPILT- - - - -GVADRSEAKVTIVGLPDIPGYAAKVFRAVAR- - - - - 286
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      245 P- - - - -PVQNRSLVGLIEIAKAVDQVADQAKVALLRVPDREGVASKLFRDIAQQQVDID 300
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      287 - - - - -RQHRHGAERLQGRQDRHLLHLLPQTSQPPPKWNTSRSTRASTQ-LLYDDHIGK 343
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      301 LIQSHDGNNDIAFTVVKDLNTAEAVTSAIAFALRSYPADQEA- - - - -EIIIVEKGIATK 357
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      344 VSLIGAMSHPGVTATFCEALAAVGNIELISTSE- - - - -DOR- - - - -SRCCAATPN 391
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      358 IAIAGAGMIGRPGIAAKMFKTLADYGVNIEMISTSEVKVSCVIDQDADRAIALSN 414
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 11

```

Q8F865 PRELIMINARY; PRT; 405 AA.
AC Q8F865;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Aspartokinase (EC 2.7.2.4).
GN LYSC OR LA0693.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RE EMBL; AEO11255; AAN47892.1; -.
KW Kinase; Transferrase; Complete proteome.
SQ SEQUENCE 405 AA; 43059 MW; DE30E782AD0CD6D9 CRC64;

```

Query Match 36.6%; Score 803; DB 16; Length 405;

```

Best Local Similarity 45.2%; Pred. No. 3.2e-46;
Matches 175; Conservative 77; Mismatches 113; Indels 22; Gaps 6;

QY      1 VQYGGSSVADAEIRIRVAERIVATKQGNVTVVVSAMGDTTDDLDAQOVCPPPPR 60
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      6 VQYGGTSVGTPERIQVAKRIKSYHDKGQVAVIVSAMGHTTDELVDLAAKISSNPKR 65
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSAGVITGTHGNNAKIIDVTPGRLOTA 120
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      66 EMDMLLTGEGQISTALLAMALWEIGVPATSETGSGIKLLTDGNFNSNAKIKMIDRSRIDA 125
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      121 LEEGRVVLVAGQGVSDTKQVTTTGRGSDTTAVMAAALGADVCEIYTDVDFGSADP 180
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      126 INEGKVVIIAGQGGIDAB-ENITTLGRGSDTSAVAVALIAGKECEIYTDVGVYTTADP 184
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 240
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      185 RVVPSAKKHQITTYEEMLELASLGAGVLHSRSLVGMNVDVVIHVRSSFNNDNIGTLVVS- 243
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      241 IKDVPWEDPILTVGAHDSKAKVTIVGLPDIPGYAAKVFRAVAR- - - - -RQHRHGA 293
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      244 -EDKINEKLVKSGVTAKNDQARITTAGVPDKPGLAAGLFGELSSKHILVDMVQSSPHNG 302
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      294 AAERLQGR-GRQDRHLLHLLPQTSQPPPKWNTSRSTRASTQ-LLYDDHIGKVLIGAGM 352
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      303 INTIGFTIPKDVH- - - - -EAKPILEGFKSHN- - - - -AKEPEINENIAIVSAVGVMK 351
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      353 SHPGVTATFCEALAAVGNIELISTSE 379
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      352 SHVGVAAAGMFKALADNGININEMISTSE 378
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```



```
RESULT 12
Q8RU02 PRELIMINARY; PRT; 408 AA.
AC Q8RU02;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Aspartate kinase (EC 2.7.2.4) (Aspartokinase).
GN LYSC.
OS Methylobacillus glycozenes.
OC Bacteria; Proteobacteria; Betaproteobacteria; Methylobacterales;
OC Methylobacillaceae; Methylobacillus.
CX NCBI_TaxID=406;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 21371;
RA Motoyama H., Anazawa H.;
RT "Analysis of the nucleotide sequence of lysC of Methylobacillus
RT glycozenes.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
CC ASPARTATE.
CC -|- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
DR EMBL; AB038354; BAB8486.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate_kinase.
DR InterPro; IPR005260; Asp_kin_monofn.
DR Pfam; PF00696; aak_kinase; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRfams; TIGR00657; asp_kinases; 1.
DR TIGRfams; TIGR00656; asp_kin_monofn; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 408 AA; 44328 MW; 3894P2315AGEBD75 CRC64;

Query Match 36.4%; Score 798.5; DB 2; Length 408;
Best Local Similarity 41.7%; Pred. No. 6.5e-46;
Matches 168; Conservative 78; Mismatches 114; Indels 43; Gaps 3;

QY 1 VQKYGSSVADAERIRVAERIVATKQGNDDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB 5 VQKYGTSVANPERIGNVARRVAEHKALGOVVVVVSAMSGEINRLIGLAKWPNPHER 64
QY 61 ELDMLLTAGERISNALVAMAEISGAHARSFTSQAGVITGTHGNAKIIDVTGRLQTA 120
DB 65 ELDVLVSTGQVTIALLSMALMELGVKAKSYTSQVSIVTDNAFTKARILKIDEEIRKD 124
QY 121 LEEGRVVLVAGFGVSDTKDVTTLGRGGSDTTAVAMAAALGADVCEIYTDVGI PSADP 180
DB 125 LDDGVVVVVTGFGQVDEH-GNITTLGRGGSDTTGVALATALKADEQIYTDVGVITTD 183
QY 181 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDREGTVVVG 240
DB 184 RVVPEARRLKSVTPEEMLEMASLGSKVLRVSVEFAGKYKVKLRVLSSPEEGEGLTIF 243
QY 241 IKQVPMEDPILTVGAHDSBAKTVIGLPIGVAAKVFAVARRR----- 286
DB 244 EDDNNMEQPIISGIAFNDEAKITVLGPDRPGIAYQILGFVDADINIDVDIIIONVAADG 303
QY 287 -----RQHRGAERLQGRGQDRHLLHLLPOTSGPPPKWNTSRSTRSASTQLLYDDH 340
DB 304 TTDFTFTVHKNDLQXALDILRDVKQGH-----GAREIIGDDK 341
QY 341 IGKVLTLGAGNRSHPGVTATFCEALAAVGNVIELISTSEDSQRS 383
DB 342 TAKVSVVGVGMRSHVGIQVSMFRTLABEGINQIMISTSEIQIS 384

RESULT 13
Q8DHWO PRELIMINARY; PRT; 599 AA.
AC Q8DHWO;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Aspartate kinase.
GN TLR1833.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
CX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AF005375; BAC09385.1; -.
KW Kinase; Complete proteome.
SQ SEQUENCE 599 AA; 62817 MW; 93D7D3ACDE2802A1 CRC64;

Query Match 35.9%; Score 786.5; DB 16; Length 599;
Best Local Similarity 45.7%; Pred. No. 7e-45;
Matches 184; Conservative 67; Mismatches 129; Indels 23; Gaps 7;

QY 1 VQKYGSSVADAERIRVAERIVATKQGNDDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB 5 VQKYGTSVGSVERIQAVRVKATVAGHQVVVVVVSAMGKTTDSLVLQVLAISDRPSQR 64
QY 61 ELDMLLTAGERISNALVAMAEISGAHARSFTSQAGVITGTHGNAKIIDVTGRLQTA 120
DB 65 EMDMLLTGEQVSTALLTMAHALGEPALSTLGAQGVITPEPAHTRARILHIETORLERH 124
QY 121 LEEGRVVLVAGFGVSDTKDVTTLGRGGSDTTAVAMAAALGADVCEIYTDVGI PSAD 179
DB 125 LRAGGVVVVAGFGGTAATDFEVTTLGRGGSDTSAVALAALRADCCIEIYTDVPGILTD 184
QY 180 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 239
DB 185 PELVFNQAQLMSEITCDEMLELASLGAKVLPRAVEIARNYGVDLVVRSSWTDDPQTRVIA 244
QY 240 -----SIKQVPMEDPILTVGAHDSBAKTVIGLPIGVAAKVFAVARRR----- 287
DB 245 PARPPRPVENVLELGRFV-DGVALDTDQAKVALLRVADRPVGAQAQFGLGARQNLVDVLI 303
QY 288 QHRHGAERLQGRGQDRHLLH--LLPOTSGPPPKWNTSRSTRSASTQLLYDDHIGKVS 345
DB 304 QSIH-----EGQTNDAFTVQKSVLKQAEALV-AFYPLSPRVEETDVLVDADIKVS 356
QY 346 LIGAGNRSHPGVTATFCEALAAVGNVIELISTSEDSQRS 388
DB 357 ITGAGMIGRPGVAAQMFSAALAAAGINLQIMISTSEVNVSVCTVAA 399

RESULT 14
Q8YR14 PRELIMINARY; PRT; 606 AA.
AC Q8YR14;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Aspartate kinase.
GN ALR3644.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
CX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
```

RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.,
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium *Anabaena* sp. strain PCC 7120.";
 RL DNA Res. 8:205-213 (2001).
 DR EMBL; AP003593; BAB75343.1; --
 DR InterPro; IPR001048; Aa_kinase.
 DR InterPro; IPR002912; ACT
 DR InterPro; IPR001341; Aspartate_kinase.
 DR InterPro; IPR005260; Asp_kin_monofn.
 DR Pfam; PF01842; ACT; 4.
 DR TIGRFAMs; TIGR00657; asp_kinases; 1.
 DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
 DR PROSITE; PS00324; ASPARTOKINASE; 1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 606 AA; 54382 MW; 932414217271D121 CRC64;

Query Match 35.4%; Score 776.5; DB 16; Length 606;
 Best Local Similarity 43.2%; Pred. No. 3.3e-44;
 Matches 175; Conservative 74; Mismatches 115; Indels 41; Gaps 7;
 QY 1 VQYGGSSVADAEIRRVRAERIVATKQGNVWVVSAMGDTTDDLLDLAQOVCAPPPR 60
 DB 5 VQKFGSSVGSVERIQVAQRYKTVKAGNSLVVVSAMGKTTDGLVKLANEISRNPNR 64
 QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTGRLQTA 120
 DB 65 EMDMLLTAGEQVITALLSNALQELGQPAISMTGQVGLVTEAETHRIILHIERVEGH 124
 QY 121 LEEGRVVLVAGFQVSQD-TKDVTTLGRGGSDTAVAMAAALGADVCEIYTDVDFGIFSAD 179
 DB 125 IKEGRVWVAGFQGISNGSLEITTLGRGGSDTSAVALAALQADFCEIYTDVDFGILT 184
 QY 180 PRIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSYDRPGTVV- 238
 DB 185 PRIVEAQLMEITCNEMLELASLGAKVLHPRAVEIARNYGNPILVVKSSWTDDEGTWTT 244
 QY 239 -----GSKDVPMDPIITGVADHRSKAKTVIVGLPDIPGVAAKVFRVARR-----R 287
 DB 245 KPQGRSLINLELARPV-DVVELTDQAKVALLRPDKPGVAAKLFGELSKQKVDVLI 303
 QY 288 QRHGAERLQGRG-----RQDRHLHLIPQTSPPPPKNTWTSRSTQ 334
 DB 304 QSIH-----EGNSNDIAFTVTTPIPKRAEAAVAIAIAPSLRSP-----SHPKSEAE 349
 QY 335 LLYDDHIGKSLIGAGMRSHPGVTATFCEALAAAGVNIELISTSE 379
 DB 350 VWVEQNIKAVSIAGAGMIGRPGVAQMFPATLAEAGVNIQIMISTSE 394

RESULT 15
 Q8G6Y6 PRELIMINARY; PRT; 254 AA.
 AC Q8G6Y6;
 DT 01-MAR-2003 (TRENDELrel. 23, Created)
 DT 01-MAR-2003 (TRENDELrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENDELrel. 23, Last annotation update)
 DE Aspartokinase.
 GN ASKA OR BL0494.
 OS Bifidobacterium longum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 OX NCBI_TaxID=216816;
 RN [1]_TaxID=216816;
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCC 2705;
 RX MEDLINE=22294977; PubMed=12381787;
 RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
 RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
 RA Pridmore R.D., Arigoni F.;

RT "The genome sequence of *Bifidobacterium longum* reflects its adaptation
 to the human gastrointestinal tract.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).
 DR EMBL; AB014669; AAN24323.1; --
 KW Kinase; Complete proteome.
 SQ SEQUENCE 254 AA; 27041 MW; 28D9F48C68822225 CRC64;
 Query Match 35.3%; Score 773.5; DB 16; Length 254;
 Best Local Similarity 64.3%; Pred. No. 1.7e-44;
 Matches 153; Conservative 38; Mismatches 46; Indels 1; Gaps 1;
 QY 1 VQYGGSSVADAEIRRVRAERIVATKQGNVWVVSAMGDTTDDLLDLAQOVCAPPPR 60
 DB 5 VQKFGSSVADTESIKRVAKRVVETEKGNKVAVVVSAMGDTTDDLLDLQALSIDSNPPR 64
 QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTGRLQTA 120
 DB 65 EMDMLLTAGERISMSLLAMAIHAESRAHSFTGQAGFTDARYGAARIKAVRPDRVXNA 124
 QY 121 LEEGRVVLVAGFQVSQD-TKDVTTLGRGGSDTAVAMAAALGADVCEIYTDVDFGIFSADP 180
 DB 125 LSLGDIIVAGFQGINA-KGDATTLGRGGSDTSAVALAALGADICEIYTDVDFGIFTADP 183
 QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSYDRPGTVV 238
 DB 184 RIVPSARRIPSIDYESILEMASCGSKVLAALRCVEYAQRNMPHVRSSFSRPPGLTV 241

Search completed: November 21, 2003, 16:08:08
 Job time : 22.8635 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 1.39184 Seconds

(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-6

Perfect score: 324

Sequence:

1 VIDLIIGTSPSWEQAAAEAV.....KITRIKLEVSFQWRPAQPR 66

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PGTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	324	100.0	66	3	US-08-818-112-77
2	324	100.0	66	4	US-08-818-111-78
3	324	100.0	66	4	US-09-056-556-77
4	324	100.0	66	4	US-09-072-596-78
5	324	100.0	802	4	US-09-056-556-214
6	324	100.0	802	4	US-09-072-596-209
7	324	100.0	802	4	US-09-072-596-346
8	69	21.3	92	4	US-09-252-991A-21917
9	61.5	19.0	657	4	US-09-300-909-16
10	58	17.9	446	4	US-09-252-991A-25318
11	57.5	17.7	500	2	US-08-031-392-7
12	57.5	17.7	500	3	US-09-299-549-7
13	57.5	17.7	500	4	US-09-610-417-7
14	57.5	17.7	633	4	US-09-252-991A-26229
15	57.5	17.7	900	4	US-09-328-352-8072
16	56	17.3	286	4	US-09-071-035-176
17	56	17.3	305	4	US-09-071-035-174
18	56	17.3	340	4	US-09-252-991A-24373
19	56	17.3	541	4	US-09-252-991A-27169
20	56	17.3	2548	4	US-09-172-422-1
21	55.5	17.1	613	4	US-09-149-727-5
22	55.5	17.1	651	4	US-09-715-858-2
23	55	17.0	597	4	US-09-252-991A-25218
24	55	17.0	1705	4	US-08-669-785-4
25	54.5	16.8	286	4	US-09-328-352-7001
26	54.5	16.8	648	4	US-09-715-858-4
27	54	16.7	360	4	US-09-252-991A-16702

28	54	16.7	450	4	US-09-252-991A-29611	Sequence 29611, A
29	53.5	16.5	363	3	US-09-046-086-2	Sequence 2, Appli
30	53.5	16.5	363	4	US-09-524-643-2	Sequence 2, Appli
31	53.5	16.5	384	4	US-09-252-991A-27463	Sequence 27463, A
32	53.5	16.5	430	4	US-09-252-991A-31388	Sequence 31388, A
33	53.5	16.5	462	1	US-08-336-408B-2	Sequence 2, Appli
34	53.5	16.5	462	3	US-08-764-870-6	Sequence 6, Appli
35	53.5	16.5	462	3	US-08-980-115-6	Sequence 6, Appli
36	53.5	16.5	462	5	PCT-US91-00399-2	Sequence 2, Appli
37	53.5	16.5	626	4	US-09-252-991A-20601	Sequence 20601, A
38	53	16.4	732	4	US-09-252-991A-19533	Sequence 19533, A
39	53	16.4	733	3	US-09-192-983-6	Sequence 6, Appli
40	53	16.4	871	4	US-09-134-001C-3979	Sequence 3979, Ap
41	53	16.4	985	4	US-09-252-991A-30698	Sequence 30698, A
42	52.5	16.2	551	2	US-09-067-351-2	Sequence 2, Appli
43	52.5	16.2	551	3	US-09-360-490-2	Sequence 2, Appli
44	52.5	16.2	745	4	US-09-107-532A-5687	Sequence 5687, Ap
45	52.5	16.2	824	4	US-09-252-991A-32329	Sequence 32329, A

ALIGNMENTS

RESULT 1

US-08-818-112-77
; Sequence 77, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED and BERY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-77

Query Match 100.0%; Score 324; DB 3; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.3e-37;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITVRIKLEVSFKM 60
DB 1 VIDIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITVRIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 2
US-08-818-111-78
; Sequence 78, Application US/0881811
; Patent No. 633852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-818-111-78
Query Match 100.0%; Score 324; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.3e-37;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITVRIKLEVSFKM 60
DB 1 VIDIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITVRIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 3
US-09-056-556-77
; Sequence 77, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 66 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-056-556-77
Query Match 100.0%; Score 324; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.3e-37;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITVRIKLEVSFKM 60
DB 1 VIDIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITVRIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 4
US-09-072-596-78
; Sequence 78, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-78
;
; Query Match 100.0%; Score 324; DB 4; Length 66;
; Best Local Similarity 100.0%; Pred. No. 2.3e-37;
; Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 VIDIIGTSPSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
; DB 1 VIDIIGTSPSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
;
; QY 61 RPAQPR 66
; DB 61 RPAQPR 66
;
; RESULT 5
; US-09-056-556-214
; Sequence 214, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
;
; Query Match 100.0%; Score 324; DB 4; Length 802;
; Best Local Similarity 100.0%; Pred. No. 5.9e-36;
; Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 VIDIIGTSPSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
; DB 9 VIDIIGTSPSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68
;
; QY 61 RPAQPR 66
; DB 61 RPAQPR 66
;
; RESULT 6
; US-09-072-596-209
; Sequence 209, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-209
;
; Query Match 100.0%; Score 324; DB 4; Length 802;
; Best Local Similarity 100.0%; Pred. No. 5.9e-36;
; Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 VIDIIGTSPSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
; DB 9 VIDIIGTSPSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68
;
; QY 61 RPAQPR 66
; DB 61 RPAQPR 66
;
; TREATM
```

Db 69 RPAQR 74

RESULT 7
US-09-072-596-346
; Sequence 346, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-072-596-346

Query Match 100.0%; Score 324; DB 4; Length 802;
Best Local Similarity 100.0%; Pred. No. 5.9e-36;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIGTSPTEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 60
|||
Db 9 VIDIGTSPTEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 68
|||

QY 61 RPAQR 66
|||
Db 69 RPAQR 74

RESULT 8
US-09-252-991A-21917
; Sequence 21917, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21917
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21917

Query Match 21.3%; Score 69; DB 4; Length 92;
Best Local Similarity 25.4%; Pred. No. 0.038;
Matches 15; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 2 IDIGTSPTEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 60
|||
Db 31 IELVGSKTSIEDAINNALAAKSIQHLWEFEVVDTRGHIENGAVGHVQVTLKVGFR 89
|||

RESULT 9
US-09-300-909-16
; Sequence 16, Application US/09300909
; Patent No. 6306580
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PREPARATION OF HUMAN PAPILLOMAVIRUS E1 HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND METHOD THEREFOR
; NUMBER OF SEQUENCES: 27
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/300,909
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/083,942
; FILING DATE: 01-MAY-1998
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-300-909-16

Query Match 19.0%; Score 61.5; DB 4; Length 657;
Best Local Similarity 29.2%; Pred. No. 5.3;
Matches 19; Conservative 14; Mismatches 29; Indels 3; Gaps 1;

QY 1 VIDIGTSPTEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITVRIKLEVS 57
|||
Db 47 MVDFITQTFCEQAELETAQALFHAQEVNDQAQVHLVKRKFAGGSGTENSPLGERLEVD 106
|||

QY 58 FQVRP 62
|||
Db 107 TELSP 111

RESULT 10
US-09-252-991A-25318
; Sequence 25318, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

QY 13 EQAAAEAVQRAR--DSVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKWR 61

CITY: BOSTON
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA: US/09/610,417
FILING DATE: 05-Jul-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/299,549
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-610-417-7

US-09-328-352-8072
; Sequence 8072, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8072
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8072

Query Match 17.7%; Score 57.5; DB 4; Length 900;
Best Local Similarity 31.7%; Pred. No. 28;
Matches 19; Conservative 9; Mismatches 21; Indels 11; Gaps 2;
Qy 13 EQAAAEAV-----QRADSVDDIRVARVIEQDMVDSAGKITRYIKLEVSFKMR 63
Db 122 EQQAREAAEQAEQLQTEQKAKATLDAAAH--QDSAAQSAFKAADVVKRGGGTVRPA 179

Search completed: November 21, 2003, 16:11:38
Job time : 2.39184 secs

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA: US/09/610,417
FILING DATE: 05-Jul-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/299,549
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-610-417-7

US-09-252-991A-26229
; Sequence 26229, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107195.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26229
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26229

Query Match 17.7%; Score 57.5; DB 4; Length 633;
Best Local Similarity 31.7%; Pred. No. 18;
Matches 20; Conservative 8; Mismatches 20; Indels 15; Gaps 3;
Qy 1 VDIICGTSTPTSEQAAAEAVQ-----RADSVDDIRVARVIEQ-DM-AVDSA 45
Db 257 VADLVGTHPDQVEAAIRVVRVENIQADQLLAHVHARPDLPDRIAAHAQAHLDMGAVDIA 316
Qy 46 GKI 48
Db 317 GAV 319

Search completed: November 21, 2003, 16:11:38
Job time : 2.39184 secs

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA: US/09/610,417
FILING DATE: 05-Jul-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/299,549
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-610-417-7

US-09-252-991A-26229
; Sequence 26229, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107195.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26229
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26229

Query Match 17.7%; Score 57.5; DB 4; Length 633;
Best Local Similarity 31.7%; Pred. No. 18;
Matches 20; Conservative 8; Mismatches 20; Indels 15; Gaps 3;
Qy 1 VDIICGTSTPTSEQAAAEAVQ-----RADSVDDIRVARVIEQ-DM-AVDSA 45
Db 257 VADLVGTHPDQVEAAIRVVRVENIQADQLLAHVHARPDLPDRIAAHAQAHLDMGAVDIA 316
Qy 46 GKI 48
Db 317 GAV 319

Search completed: November 21, 2003, 16:11:38
Job time : 2.39184 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 3.99592 seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-6

Perfect score: 324

Sequence: 1 VIDIIQTSTSEWQAAEAV.....KITRIKLEVSFMRPAQR 66

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 segs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq 19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	324	100.0	66	18	AAW32433
2	324	100.0	66	18	AAW32265
3	324	100.0	66	19	AAW81668
4	324	100.0	66	19	AAW64305
5	324	100.0	66	20	AAW39107
6	324	100.0	66	20	AAW38970
7	324	100.0	66	23	AAW29718
8	324	100.0	66	23	AAW17582
9	324	100.0	802	19	AAW81746

10	324	100.0	802	19	AAW64379	Mycobacterium anti
11	324	100.0	802	20	AAW32063	Mycobacterium tube
12	324	100.0	802	20	AAW32224	M. tuberculosis fu
13	324	100.0	802	20	AAW39176	M. tuberculosis fu
14	324	100.0	802	20	AAW39081	M. tuberculosis fu
15	324	100.0	802	20	AAW39033	M. tuberculosis fu
16	324	100.0	802	23	AAU74592	Antigenic fusion p
17	324	100.0	983	22	AAU01901	M. tuberculosis Tb
18	61.5	19.0	253	12	AAW13507	P.denitrificans CO
19	61.5	19.0	368	22	ABB69744	Drosophila melanog
20	61.5	19.0	657	21	AAW57582	Human papillomavir
21	61.5	19.0	657	22	AAW86424	Human papillomavir
22	61.5	19.0	1477	22	ABB61960	Drosophila melanog
23	61	18.8	330	22	AAU38359	Salmonella typhi c
24	60	18.5	224	23	ABP31340	Human ORF313 prote
25	60	18.5	535	24	ABU08348	Cotton (+)-delta-c
26	60	18.5	1624	23	ABU05888	M. tuberculosis an
27	59.5	18.4	1938	24	ABP76679	Streptomyces virid
28	59	18.2	410	23	AAE17876	Sucrose-specific p
29	59	18.2	984	22	ABB61888	Drosophila melanog
30	59	18.2	1221	23	AAE34373	Fruit fly E93 prog
31	58.5	18.1	324	21	AAW39229	Gene 2 human secre
32	58.5	18.1	324	21	AAW39230	Human secreted pro
33	58.5	18.1	398	21	AAW43819	Human cancer assoc
34	58.5	18.1	398	23	ABP41571	Human ovarian anti
35	58.5	18.1	1005	21	AAW31946	Human cytoskeleton
36	58	17.9	110	21	AAW5654	Arabidopsis thalia
37	58	17.9	191	21	AAW24668	Plant SDF encoded
38	58	17.9	617	22	AAW25464	Human mdt protein
39	58	17.9	974	19	AAW55960	Human transient re
40	58	17.9	1737	23	AAO14211	Human transporter
41	58	17.9	1738	23	ABB8491	Human Canlon prote
42	58	17.9	1738	24	ABP96139	Human cation chann
43	58	17.9	1738	24	ABP57388	Human BION1 protei
44	57.5	17.7	255	13	AAW21516	Polyfunctional pro
45	57.5	17.7	441	20	AAW40501	A. thaliana yael p

ALIGNMENTS

RESULT 1	AAW32433	ID	AAW32433 standard; Protein; 66 AA.
XX	AC	AAW32433;	
XX	DT	06-JAN-1998 (first entry)	
XX	DE	Mycobacterium tuberculosis antigen TBra3.	
XX	KW	Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;	
XX	KW	Skin testing; M.tuberculosis.	
XX	OS	Mycobacterium tuberculosis.	
XX	PN	WO9709428-A2.	
XX	PD	13-MAR-1997.	
XX	PF	30-AUG-1996; 96WO-US14674.	
XX	PR	12-JUL-1996; 96US-0680574.	
XX	PR	01-SEP-1995; 95US-0523436.	
XX	PR	22-SEP-1995; 95US-0533634.	
XX	PR	22-MAR-1996; 96US-0620874.	
XX	PR	05-JUN-1996; 96US-0659683.	
XX	PA	(CORI-) CORIXA CORP.	
XX	PI	Campos-neto A. Dillon DC, Houghton R, Reed SG, Skeiky YAW;	
XX	PI	Twardzik DR, Vedwick TH,	

DR WPI; 1997-192903/17.
 DR N-PSDB; AAT91475.
 XX
 PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also
 PT for diagnosis
 PT
 XX
 PS Example 3; Page 113; 169pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, TbrA3. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX
 SQ Sequence 66 AA;
 Query Match 100.0%; Score 324; DB 18; Length 66;
 Best Local Similarity 100.0%; Pred. No. 1e-35;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VIDIIGTPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
 DB 1 VIDIIGTPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
 QY 61 RPAQPR 66
 DB 61 RPAQPR 66
 RESULT 2
 AAW32365
 ID AAW32365 standard; Protein; 66 AA.
 XX
 AC AAW32365;
 XX
 DT 13-JAN-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen TbrA3.
 XX
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9709429-A2.
 XX
 PD 13-MAR-1997.
 XX
 PF 30-AUG-1996; 96WO-US14675.
 XX
 PR 12-JUL-1996; 96US-0680573.
 PR 01-SEP-1995; 95US-0523435.
 PR 22-SEP-1995; 95US-0532136.
 PR 22-MAR-1996; 96US-0620280.
 PR 05-JUN-1996; 96US-0658800.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
 PI Twardzik DR, Vedvick TH;
 XX
 WPI; 1997-192904/17.
 DR N-PSDB; AAT91412.
 XX
 PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
 PT - useful for diagnosis of M. tuberculosis infection
 XX
 PS Example 3; Page 123; 190pp; English.

XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, TbrA3. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX
 SQ Sequence 66 AA;
 Query Match 100.0%; Score 324; DB 18; Length 66;
 Best Local Similarity 100.0%; Pred. No. 1e-35;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VIDIIGTPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
 DB 1 VIDIIGTPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
 QY 61 RPAQPR 66
 DB 61 RPAQPR 66
 RESULT 3
 AAW81668
 ID AAW81668 standard; Protein; 66 AA.
 XX
 AC AAW81668;
 XX
 DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis immunogenic polypeptide TbrA3.
 XX
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US18293.
 XX
 PR 13-MAR-1997; 97US-0818112.
 PR 11-OCT-1996; 96US-0730510.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 WPI; 1998-261042/23.
 DR N-PSDB; AAV64461.
 XX
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX
 PS Example 3; Page 109-110; 230pp; English.
 XX
 CC This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
 CC for inducing protective immunity against tuberculosis (TB). This sequence
 CC can be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis.
 XX
 SQ Sequence 66 AA;

```
Query Match          100.0%; Score 324; DB 19; Length 66;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTPTSWEQAAAEAVQARQSDVDDIRVARVIEQDMAYDSAGKITRYIKLEVSPFM 60
DB 1 VIDIIGTPTSWEQAAAEAVQARQSDVDDIRVARVIEQDMAYDSAGKITRYIKLEVSPFM 60

QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 5
AAV39107
ID AAV39107 standard; Protein; 66 AA.
XX
AC AAV39107;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis antigen Tbra3 amino acid sequence.
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US03268.
XX
PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527409/44.
DR N-PSDB; AAV44353.
XX
PT New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX
PS Example 3; Page 107; 299pp; English.
XX
CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAV39083 to
CC AAV39225 are used in the exemplification of the present invention.
XX
SQ Sequence 66 AA;

Query Match          100.0%; Score 324; DB 20; Length 66;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTPTSWEQAAAEAVQARQSDVDDIRVARVIEQDMAYDSAGKITRYIKLEVSPFM 60
DB 1 VIDIIGTPTSWEQAAAEAVQARQSDVDDIRVARVIEQDMAYDSAGKITRYIKLEVSPFM 60

QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 4
AAW64305
ID AAW64305 standard; Protein; 66 AA.
XX
AC AAW64305;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen Tbra3.
XX
KW Tuberculosis; infection; diagnosis; antigen; Tbra3.
XX
OS Mycobacterium tuberculosis strain H37Ra.
XX
PN WO9816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-251292/22.
DR N-PSDB; AAV44353.
XX
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Example 3; Page 114; 250pp; English.
XX
CC This polypeptide comprises Mycobacterium tuberculosis soluble
CC antigen Tbra3. It is encoded by a DNA sequence (see AAV44353)
CC isolated from a M. tuberculosis strain H37Ra expression library
CC with rabbit anti-sera raised against M. tuberculosis supernatant.
CC No significant homology was found between Tbra3 and Genebank
CC database sequences. The invention relates to compositions and
CC methods for diagnosing tuberculosis. It provides polypeptides
CC (see AAW64291-W64379) comprising an antigenic portion of a soluble
CC M. tuberculosis antigen, or an immunogenic portion of an M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using the
CC above polypeptides, antibodies or oligonucleotide probes and
CC primers, for the diagnosis of tuberculosis.
XX
SQ Sequence 66 AA;

Query Match          100.0%; Score 324; DB 19; Length 66;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RESULT 6
 AAY38970
 ID AAY38970 standard; Protein; 66 AA.
 XX
 AC AAY38970;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis recombinant antigen protein TBra3.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 OS
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03265.
 XX
 PR 05-MAY-1998; 98US-0072596.
 PR 18-FEB-1998; 98US-0024753.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 XX WPI; 1999-527416/44.
 DR N-PSDB; AA219051.
 XX
 XX New polypeptide comprising antigenic portions of M. tuberculosis
 PS Example 3; Page 151; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX
 SQ Sequence 66 AA;
 Query Match 100.0%; Score 324; DB 20; Length 66;
 Best Local Similarity 100.0%; Pred. No. 1e-35;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VIDIIGTSTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITTYRIKLEVSFKM 60
 DB 1 VIDIIGTSTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITTYRIKLEVSFKM 60
 QY 61 RPAQPR 66
 DB 61 RPAQPR 66
 RESULT 7
 AAE29718
 ID AAE29718 standard; Protein; 66 AA.
 XX
 AC AAE29718;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium tuberculosis TBra3 antigenic protein.
 KW Vaccine; immunity; diagnostic agent; gene therapy; TBra3 antigen.
 XX
 OS Mycobacterium tuberculosis.

XX WO200272792-A2.
 PN 19-SEP-2002.
 XX
 PF 13-MAR-2002; 2002WO-US08223.
 XX
 PR 13-MAR-2001; 2001US-275837P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Brannon M, Guderian J;
 XX WPI; 2002-759844/82.
 DR N-PSDB; AA47095.
 XX
 PT New recombinant nucleic acid molecule comprising a Leishmania TSA,
 PT Leif, M15 or 6H polynucleotide, useful as vaccine to elicit protective
 PT immunity against pathogenic microorganisms e.g. Leishmania and
 PT Mycobacterium tuberculosis
 XX
 PS Disclosure; Page 107-108; 155pp; English.
 XX
 CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides,
 CC as in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC M. tuberculosis TBra3 antigenic protein.
 XX
 SQ Sequence 66 AA;
 Query Match 100.0%; Score 324; DB 23; Length 66;
 Best Local Similarity 100.0%; Pred. No. 1e-35;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VIDIIGTSTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITTYRIKLEVSFKM 60
 DB 1 VIDIIGTSTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITTYRIKLEVSFKM 60
 QY 61 RPAQPR 66
 DB 61 RPAQPR 66
 RESULT 8
 AAE17582
 ID AAE17582 standard; Protein; 66 AA.
 XX
 AC AAE17582;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium species TBra3 protein.
 XX
 KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; TBra3 protein.
 XX
 OS Mycobacterium sp.
 XX
 PN WO200198460-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 20-JUN-2001; 2001WO-US19959.
 XX

PF 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; -97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Iwardzik DR, Vedvick IS;
PI N-PSDB; AAV64567.
XX
DR WPI; 1998-261042/23.
DR N-PSDB; AAV64567.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
XX
XX Disclosure; Page 208-211; 230pp; English.
PS
CC This sequence represents the fusion protein Tbf-2 which is composed of
CC immunogenic polypeptides from Mycobacterium tuberculosis (MT). This
CC protein is used in a method for inducing protective immunity against
CC tuberculosis (TB). This sequence can be formulated into vaccines
CC and/or pharmaceutical compositions for immunising against
CC M. tuberculosis infection or may be used for the diagnosis of TB.
XX
SQ Sequence 802 AA;

Query Match 100.0%; Score 324; DB 19; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34; Indels 0; Gaps 0;
Matches 66; Conservative 0; Mismatches 0

QY 1 VIDITGTSPTSWEQAAAEVQRARDSDIRVARVIEQDMAVDASAGKITRIKLEVSFKM 60
Db 9 VIDITGTSPTSWEQAAAEVQRARDSDIRVARVIEQDMAVDASAGKITRIKLEVSFKM 68
QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 10
AAW64379
ID AAW64379 standard; Protein; 802 AA.
XX
AC AAW64379;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium antigen Tbf2 protein fusion.
XX
KW Tuberculosis; infection; diagnosis; 38 kDa antigen; Tbra3; DPEP;
KW Tbf38-1; Tbf-2.
XX
OS Mycobacterium tuberculosis.
OS Synthetic.
XX
PN WO9816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Iwardzik DR, Vedvick IS;
PI N-PSDB; AAV55801.
XX
DR WPI; 1998-251232/22.
DR N-PSDB; AAV55801.

PF 20-JUN-2000; 2000US-0597796.
XX
PR 01-FEB-2001; 2001US-265737P.
XX
PA (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Alderson M;
XX
XX WPI; 2002-147798/19.
XX
XX N-PSDB; AAD28353.
XX
PT Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a
PT subject -
XX
XX Claim 9; Page 123-124; 136pp; English.
PS
CC The present invention relates to fusion proteins containing at least
CC two Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC Polynucleotides are useful as diagnostic tools in patients infected
CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines, MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC Tbra3 protein.
XX
XX
SQ Sequence 66 AA;

Query Match 100.0%; Score 324; DB 23; Length 66;
Best Local Similarity 100.0%; Pred. No. 1e-35; Indels 0; Gaps 0;
Matches 66; Conservative 0; Mismatches 0

QY 1 VIDITGTSPTSWEQAAAEVQRARDSDIRVARVIEQDMAVDASAGKITRIKLEVSFKM 60
Db 1 VIDITGTSPTSWEQAAAEVQRARDSDIRVARVIEQDMAVDASAGKITRIKLEVSFKM 60
QY 61 RPAQPR 66
Db 61 RPAQPR 66

RESULT 9
AAW81746
ID AAW81746 standard; Protein; 802 AA.
XX
AC AAW81746;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis fusion protein Tbf-2.
XX
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
XX Synthetic.
XX OS Mycobacterium tuberculosis.
XX
XX WO9816646-A2.
XX
XX 23-APR-1998.
XX
XX PD

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX Example 7; Page 223-226; 250pp; English.

XX This polypeptide comprises a fusion protein, designated Tbf-2,
CC composed of Mycobacterium tuberculosis antigens TbrA3 (see AAW64295),
CC 38 kDa antigen (see AAW64364), Tb38-1 (see AAW64321) and DPEP (see
CC AAW64322). It was produced by PCR amplification (see AAW44450-57) of
CC the appropriate antigen DNA sequences, cloning into an expression
CC vector, and expression in E. coli. Tbf-2 can be used for
CC serodiagnosis of tuberculosis. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising antigenic or
CC immunogenic portions of M. tuberculosis antigens, or fusion proteins,
CC DNA sequences encoding such polypeptides, recombinant expression
CC vectors and host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient.

XX Sequence 802 AA;
SQ Query Match 100.0%; Score 324; DB 19; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
DB 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68
QY 61 RPAQPR 66
DB 69 RPAQPR 74

RESULT 11
AAV32063
ID AAY32063 standard; Protein; 802 AA.
AC AAY32063;
XX 17-JAN-2000 (first entry)
DE Mycobacterium tuberculosis antigen fusion protein Tbf-2.
XX Tuberculosis; antigen; fusion protein; Tbf-2; TbrA3; 38kD; Tb38-1;
KW DPEP; diagnosis; therapy; vaccine; immunogen.
XX Mycobacterium tuberculosis.
OS WO9951748-A2.
PN WO9951748-A2.
XX 14-OCT-1999.
PD 07-APR-1999; 99WO-US07717.
XX 07-APR-1999; 98US-0056556.
PR 30-DEC-1998; 98US-0223040.
XX (CORI-) CORIXA CORP.
PA Skeiky YAW, Alderson M, Campos-Neto A;
XX WPI; 1999-601610/51.
DR N-PSDB; AAZ20198.
XX New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -
XX Claim 1; Fig 5G-J; 83pp; English.
PS This sequence represents a recombinant Mycobacterium tuberculosis
CC

CC tetra-antigen fusion protein, termed Tbf-2, composed of the antigens
CC TbrA3, 39kD, Tb38-1 and DPEP. The fusion protein is expressed in
CC host cells using a vector carrying a polynucleotide (see AAZ20198)
CC comprising the 4 coding sequences. The invention provides fusion
CC proteins (see AAY32059-71) containing at least 2 M. tuberculosis
CC antigens. The new fusion proteins and polynucleotides encoding
CC them are useful as vaccines for preventing tuberculosis (claimed),
CC for diagnosis (via in vitro assays or intradermal skin tests for
CC detection of anti-M. tuberculosis antibodies), monitoring of
CC disease progression, and treatment of tuberculosis. They are more
CC effective immunogens than mixtures of the individual protein
CC components.

XX Sequence 802 AA;
SQ Query Match 100.0%; Score 324; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
DB 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68
QY 61 RPAQPR 66
DB 69 RPAQPR 74

RESULT 12
AAV39224
ID AAY39224 standard; Protein; 802 AA.
XX AAY39224;
AC AAY39224;
XX 05-NOV-1999 (first entry)
DT M. tuberculosis fusion protein Tbf-6 amino acid sequence.
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX Synthetic.
OS Mycobacterium tuberculosis.
XX WO9942076-A2.
PN 26-AUG-1999.
PD 17-FEB-1999; 99WO-US03268.
XX 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527409/44.
DR N-PSDB; AAZ19457.
XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX Claim 37; Page 271-273; 299pp; English.
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to

CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 802 AA;

Query Match 100.0%; Score 324; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIICTSPSTWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITIRIKLEVSFKM 60
DB 9 VIDIICTSPSTWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITIRIKLEVSFKM 68

QY 61 RPAQPR 66
DB 69 RPAQPR 74

RESULT 13
AAY39176
ID AAY39176 standard; Protein; 802 AA.
XX
AC AAY39176;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis fusion protein TbF-2 amino acid sequence.
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
OS Synthetic.
CS Mycobacterium tuberculosis.
XX
FN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US03268.
XX
PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R,
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527409/44.
DR N-PSDB; AA219368.
XX
PT New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX
PS Disclosure; Page 205-208; 299pp; English.

CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 802 AA;

Query Match 100.0%; Score 324; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIICTSPSTWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITIRIKLEVSFKM 60
DB 9 VIDIICTSPSTWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITIRIKLEVSFKM 68

QY 61 RPAQPR 66
DB 69 RPAQPR 74

RESULT 14
AAY39081
ID AAY39081 standard; Protein; 802 AA.
XX
AC AAY39081;
XX
DT 05-NOV-1999 (first entry)
XX
DE M tuberculosis fusion protein TbF-6.
XX
KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX
OS Mycobacterium tuberculosis.
XX
FN WO9942118-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US03265.
XX
PR 05-MAY-1998; 98US-0072596.
PR 18-FEB-1998; 98US-0024753.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R,
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527416/44.
DR N-PSDB; AA219245.
XX
PT New polypeptide comprising antigenic portions of M. tuberculosis
PS Example 10; Page 316-318; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX
SQ Sequence 802 AA;

Query Match 100.0%; Score 324; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIICTSPSTWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITIRIKLEVSFKM 60
DB 9 VIDIICTSPSTWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITIRIKLEVSFKM 68

QY 61 RPAQPR 66
DB 69 RPAQPR 74

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 2.57415 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-6
Perfect score: 324
Sequence: 1 VIDIGSTPSNEQAAAEV.....KITYRIKLEVSFKMRPAQPR 66

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559496 residues
Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	324	100.0	66	12	US-10-084-843-77	Sequence 77, Appl
2	324	100.0	66	12	US-10-193-002-78	Sequence 78, Appl
3	324	100.0	66	12	US-10-098-732A-37	Sequence 37, Appl
4	324	100.0	802	9	US-09-287-849-10	Sequence 10, Appl
5	324	100.0	802	12	US-10-084-843-214	Sequence 214, App
6	324	100.0	802	12	US-10-084-843-351	Sequence 351, App
7	324	100.0	802	12	US-10-193-002-209	Sequence 209, App
8	324	100.0	802	12	US-10-193-002-346	Sequence 346, App
9	324	100.0	802	12	US-10-359-460-10	Sequence 10, Appl
10	63.5	19.6	342	12	US-10-166-225A-63	Sequence 63, Appl
11	62.5	19.3	561	15	US-10-078-770-154	Sequence 154, App
12	61.5	19.0	1073	16	US-10-210-115-20	Sequence 20, Appl
13	61	18.8	330	9	US-09-815-242-13952	Sequence 13952, A
14	60	18.5	535	14	US-10-067-534-2	Sequence 2, Appl
15	60	18.5	1624	16	US-10-080-170-539	Sequence 539, App

15	59	18.2	410	10	US-09-905-176-26	Sequence 26, Appl
16	59	18.2	1221	14	US-10-016-768-11	Sequence 11, Appl
17	58.5	18.1	398	9	US-09-925-301-1264	Sequence 1264, Appl
18	58	17.9	828	15	US-10-198-070-98	Sequence 98, Appl
19	58	17.9	836	15	US-10-198-070-97	Sequence 97, Appl
20	58	17.9	974	12	US-10-153-244-106	Sequence 106, Appl
21	58	17.9	974	12	US-10-210-152-23	Sequence 23, Appl
22	58	17.9	977	15	US-10-198-070-108	Sequence 108, Appl
23	58	17.9	982	15	US-10-198-070-63	Sequence 63, Appl
24	58	17.9	982	15	US-10-198-070-72	Sequence 72, Appl
25	58	17.9	1738	15	US-10-178-791-1	Sequence 1, Appl
26	58	17.9	1738	15	US-10-178-791-3	Sequence 3, Appl
27	58	17.9	500	10	US-09-981-947A-7	Sequence 7, Appl
28	57.5	17.7	861	15	US-10-156-761-11328	Sequence 11328, A
29	57.5	17.7	323	15	US-10-198-070-64	Sequence 64, Appl
30	57	17.6	1115	10	US-09-866-582-40	Sequence 40, Appl
31	56.5	17.4	94	12	US-10-029-386-28176	Sequence 28176, A
32	56	17.3	515	12	US-10-306-905-12	Sequence 12, Appl
33	56	17.3	2548	9	US-09-851-682A-1	Sequence 1, Appl
34	56	17.3	460	16	US-10-278-536-2	Sequence 2, Appl
35	55.5	17.1	489	10	US-09-826-581-6	Sequence 6, Appl
36	55.5	17.1	613	12	US-10-120-145-5	Sequence 5, Appl
37	55.5	17.1	722	15	US-10-136-841-6	Sequence 6, Appl
38	55.5	17.1	85	15	US-10-156-761-14857	Sequence 14857, A
39	55	17.0	888	15	US-10-097-340-161	Sequence 161, Appl
40	55	17.0	928	9	US-09-801-574-44	Sequence 44, Appl
41	55	16.8	310	15	US-10-156-761-12054	Sequence 12054, A
42	54.5	16.8	1107	15	US-10-153-219-4	Sequence 4, Appl
43	54.5	16.8	1143	15	US-10-153-219-6	Sequence 6, Appl
44	54.5	16.8	1143	15	US-10-153-219-8	Sequence 8, Appl
45	54.5	16.8	1143	15	US-10-153-219-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-10-084-843-77
; Sequence 77, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.

```
/
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.411C9
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 77:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 66 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-084-843-77

Query Match 100.0%; Score 324; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.7e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
Db 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60

QY 61 RPAQPR 66
Db 61 RPAQPR 66

RESULT 2
US-10-193-002-78
; Sequence 78, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 66 amino acids
```

```
/
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-10-193-002-78

Query Match 100.0%; Score 324; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.7e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
Db 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60

QY 61 RPAQPR 66
Db 61 RPAQPR 66

RESULT 3
US-10-098-732A-37
; Sequence 37, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Tbra3
US-10-098-732A-37

Query Match 100.0%; Score 324; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.7e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
Db 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60

QY 61 RPAQPR 66
Db 61 RPAQPR 66

RESULT 4
US-09-287-849-10
; Sequence 10, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
```

;; CURRENT FILING DATE: 1999-04-07
;; PRIOR APPLICATION NUMBER: US 08/818,112
;; PRIOR FILING DATE: 1997-03-13
;; PRIOR APPLICATION NUMBER: US 08/942,578
;; PRIOR FILING DATE: 1997-10-01
;; PRIOR APPLICATION NUMBER: US 09/025,197
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 09/056,556
;; PRIOR FILING DATE: 1998-04-07
;; PRIOR APPLICATION NUMBER: US 09/223,040
;; PRIOR FILING DATE: 1998-12-30
;; NUMBER OF SEQ ID NOS: 46
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 10
;; LENGTH: 802
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-10

Query Match 100.0%; Score 324; DB 9; Length 802;
Best Local Similarity 100.0%; Pred.No.3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDVAVDSAGKITVRIKLEVSFKM 60
Db 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDVAVDSAGKITVRIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 5
US-10-084-843-214
; Sequence 214, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392

;; REFERENCE/DOCKET NUMBER: 210121.411C9
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 214:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 802 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; SEQUENCE DESCRIPTION: SEQ ID NO: 214:
US-10-084-843-214

Query Match 100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred.No.3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDVAVDSAGKITVRIKLEVSFKM 60
Db 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDVAVDSAGKITVRIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 6
US-10-084-843-351
; Sequence 351, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 351:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 802 amino acids
;; TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 351:
US-10-084-843-351

Query Match      100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 7
US-10-193-002-209
; Sequence 209, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-4900
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 209:
US-10-193-002-209

Query Match      100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 8
US-10-193-002-346
; Sequence 346, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 346:
US-10-193-002-346

Query Match      100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
```

```
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 8
US-10-193-002-346
; Sequence 346, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 346:
US-10-193-002-346

Query Match      100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
```

```
Db          |||||
69 RPAQPR 74

RESULT 9
US-10-359-460-10
; Sequence 10, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dallon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tetra-fusion
US-10-359-460-10

Query Match          100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 VIDIIGTSWEQAAAEAVQARSDSVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 60
Db          |||||
9 VIDIIGTSWEQAAAEAVQARSDSVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 68

Qy          61 RPAQPR 66
Db          |||||
69 RPAQPR 74

RESULT 10
US-10-166-225A-63
; Sequence 63, Application US/10166225A
; Publication No. US20030148416A1
; GENERAL INFORMATION:
; APPLICANT: BEREY, Alan
; APPLICANT: BRETZEL, Werner
; APPLICANT: HUMSELIN, Markus
; APPLICANT: LOPEZ-ULIBARRI, Rual
; APPLICANT: MAYER, Anne F.
; APPLICANT: YELISSEV, Alexei A.
; TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION
; FILE REFERENCE: C38435/121966
; CURRENT APPLICATION NUMBER: US/10/166,225A
; CURRENT FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 63

Db          |||||
69 RPAQPR 74

Query Match          19.3%; Score 62.5; DB 15; Length 561;
Best Local Similarity 32.1%; Pred. No. 12;
Matches 18; Conservative 12; Mismatches 23; Indels 3; Gaps 2;

Qy          14 QAAAEAVQARSDSVDDIRVARVIEQDM--AVDSAGKITVRI-KLEVSFKMREPAQR 66
Db          |||||
228 EAAKVLVKEKGDVVAQLAAKAKSKQEIETAVDELNRAKEIVSKLEERFKLKEGIPR 283

RESULT 12
US-10-210-115-20
; Sequence 20, Application US/10210115
; Publication No. US20030129708A1
; GENERAL INFORMATION:
; APPLICANT: Puitsyn, Leonid R
; APPLICANT: Smirnov, Sergey V
; APPLICANT: Altman, Irina B
; APPLICANT: No. US20030129708A1Ikovva, Anna E
; APPLICANT: Kotliarova, Veronika A
```


Search completed: November 21, 2003, 16:38:11
Job time : 2.57415 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 1.37687 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-6
Perfect score: 324
Sequence: 1 VIDIICTSPTSWEQAAAEAV.....KITYRIKLEVSFWMRPAQPR 66

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: Piri:*
2: Piri2:*
3: Piri3:*
4: Piri4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	324	100.0	71	2 F70833	probable protein t
2	101	31.2	77	2 E84298	hypothetical prote
3	69	21.3	71	2 E83641	hypothetical prote
4	66	20.4	69	2 D95850	hypothetical prote
5	65.5	20.2	1073	2 D90633	carbamoyl-phosphat
6	65.5	20.2	1073	2 C85494	carbamoyl-phosphat
7	64	19.8	96	2 A21822	transcription regu
8	63.5	19.6	342	2 C71704	hypothetical prote
9	63.5	19.6	342	2 H97792	carotenoid biosynt
10	62	19.1	282	2 AD3528	flagellin (impor
11	62	19.1	1094	2 S49313	protein kinase - s
12	61.5	19.0	253	2 H36145	El protein - Pse
13	61.5	19.0	657	1 W1WL18	Beta-glucosidase
14	61.5	19.0	947	2 S08243	beta-glucosidase
15	61.5	19.0	1073	1 SYECP	carbamoyl-phosphat
16	61.5	19.0	1477	2 T13797	tumor supressor pr
17	61	18.8	94	2 E84227	hypothetical prote
18	61	18.8	2626	2 T31099	myosin-RhoGAP prot
19	60.5	18.7	1075	2 A20510	carbamoyl-phosphat
20	60	18.5	181	2 S23604	hypothetical prote
21	60	18.5	246	2 A22652	cell division prot
22	60	18.5	267	2 C97434	hypothetical prote
23	60	18.5	396	2 H98803	A3C transporter (A
24	60	18.5	1624	2 C70867	Probable Helix-tur
25	59.5	18.4	458	2 A43554	desmin - African c
26	59.5	18.4	620	2 H69382	ABC transporter, A
27	59.5	18.4	959	2 H72699	probable leucyl-tr
28	59	18.2	393	2 AH2916	acetyl-CoA acetyl
29	59	18.2	393	2 C97691	chain A, unligande

ALIGNMENTS

RESULT 1

F70833
Probable protein transport protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70833
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295387; PMID: 9634230
A: Accession: F70833
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-71 <COL>
A: Cross-references: GB:AL021931; GB:AL123456; NID:G3261526; PIDN:CAA17385.1; PID:ei252494
A: Experimental source: strain H37RV
C: Geneticks:
A: Gene: sec

Query Match 100.0%; Score 324; DB 2; Length 71;
Best Local Similarity 100.0%; Pred. No. 5e-30;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIICTSPTSWEQAAAEAVQARDSDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
DB 6 VIDIICTSPTSWEQAAAEAVQARDSDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 65
QY 61 RPAQPR 66
DB 66 RPAQPR 71
RESULT 2
E84298
hypothetical protein Vng1446h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84298
R: Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A: Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A: Title: Genome sequence of Halobacterium species NRC-1
A: Reference number: A84160; MUID: 20504483; PMID: 11016950
A: Accession: E84298
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-77 <STO>


```
Matches 19; Conservative 17; Mismatches 19; Indels 7; Gaps 2;
QY 2 IDIIGTSPTSWEQAAAEAVQARDSDVDIRV-----ARVIEQDMAVDSAGKITVRIKLEV 56
Db 659 VPVIGTSPDAIR--AEDRRFQHAVDRUKLKQAPANATVTAIEMAVEKAEIGYPLVWRP 716
QY 57 SF 58
Db 717 SY 718

RESULT 7
AE1822
transcription regulator asl0125 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AE1822
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE1822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-96 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA577649.1; PID:gl7135103; GSPDB:GN00179
A:Experimental source: Strain PCC 7120
C:Genetics:
A:Gene: asl0125

Query Match 19.8%; Score 64; DB 2; Length 96;
Best Local Similarity 26.7%; Pred. No. 2.8;
Matches 16; Conservative 10; Mismatches 28; Indels 6; Gaps 1;
QY 7 TSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITVRIKLEVSFMRPAQPR 66
Db 26 TSVRNWEAGRTPTPKYVDALRIFKCTPEELASASEKSINQR-----HKRKEGRPR 79

RESULT 8
C71704
hypothetical protein RP452 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: C71704
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: C71704
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-342 <AND>
A:Cross-references: GB:AJ235271; GB:AJ235269; MID:g3868717; PIDN:CAA14909.1; PID:g386100
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: RP452
C:Superfamily: carotenoid biosynthesis protein homolog

Query Match 19.6%; Score 63.5; DB 2; Length 342;
Best Local Similarity 25.8%; Pred. No. 13;
Matches 16; Conservative 14; Mismatches 13; Indels 19; Gaps 2;
QY 1 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITVRIK 44
Db 215 VLDIAGSGGTSWSQVEA---YRATNSLQNRASSFINWGIIPTLDSLKMVREVSQDPIIT 271

RESULT 9
H97792
carotenoid biosynthesis proteins homolog [imported] - Rickettsia conorii (strain Malish ;
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: H97792
R;Cgata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: H97792
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-342 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03282.1; PID:gl5619839; GSPDB:GN00173
C:Genetics:
A:Gene: RC0744
C:Superfamily: carotenoid biosynthesis protein homolog

Query Match 19.6%; Score 63.5; DB 2; Length 342;
Best Local Similarity 25.8%; Pred. No. 13;
Matches 16; Conservative 13; Mismatches 14; Indels 19; Gaps 2;
QY 1 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDS 44
Db 215 VLDIAGSGGTSWSQVEA---YRATNSLQNRASSFINWGIIPTLDSLKMVREVSQDPIIT 271
QY 45 AG 46
Db 272 SG 273

RESULT 10
AD3528
flagellin [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AD3528
R;DelVecchio, V.G.; Kapurral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis;
A:Reference number: AD3528; PMID:11756688
A:Accession: AD3528
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-282 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL53391.1; PID:gl7984285; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI10150
A:Map position: II

Query Match 19.1%; Score 62; DB 2; Length 282;
Best Local Similarity 36.8%; Pred. No. 15;
Matches 14; Conservative 6; Mismatches 18; Indels 0; Gaps 0;
QY 16 AAEEAVQARDSDVDIRVARVIEQDMAVDSAGKITVRIK 53
Db 77 AYSAINKIRSVDDIKTKLVSMANGASTEDKGIETIK 114

RESULT 11
S49313
protein kinase - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C>Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 24-Sep-1999
C:Accession: S52076; S49313
R;Wetterauer, B.W.; Hamker, U.; von Haeseler, A.; MacWilliams, H.K.; Simon, M.N.; Veron,
Biochim. Biophys. Acta 1285, 97-101, 1995
A:Title: A protein kinase from Dictyostelium discoideum with an unusual acidic repeat don
A:Reference number: S52076; MUID:95161460; PMID:7857991
```

A;Accession: S52076
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1094 <WE2>
A;Cross-references: EMBL:Z37981; NID:G551445; PIDN:CAA86053.1; PID:G551446
C;Genetics:
A;Introns: 35/3; 104/1; 166/2
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
F:20-281/Domain: protein kinase homology <KIN>
C;Keywords: ATP
F:28-36/Region: protein kinase ATP-binding motif

Query Match 19.1%; Score 62; DB 2; Length 1094;
Best Local Similarity 30.0%; Pred. No. 68;
Matches 15; Conservative 13; Mismatches 20; Indels 2; Gaps 1;

QY 13 EQAAAEAVQARDSDVDIRVARVI--EODMAVDSAGKITRYIKLEVSFKM 60
Db 893 EVKAEPTKAESEVDKVEPIKVEPVPKVEEPKVEEPKVEPIKV 942

RESULT 12
H36145
cobM protein - Pseudomonas sp.
C;Species: Pseudomonas sp.
C;Date: 30-Nov-1990 #sequence_revision 30-Nov-1990 #text_change 24-Sep-1999
C;Accession: H36145
J. Bacteriol. 172, 5980-5990, 1990
R. Crouzet, J.; Cameron, B.; Cauchots, L.; Rigault, S.; Rouyez, M.C.; Blanche, F.; Thibaut
A;Title: Genetic and sequence analysis of an 8.7-kilobase Pseudomonas denitrificans frag
A;Reference number: A36145; MUID:91008976; PMID:2211521
A;Accession: H36145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <CRO>
A;Cross-references: GB:M59301; GB:M32224; NID:G151170; PIDN:AAA25801.1; PID:G151178
A;Note: the source is designated as Pseudomonas denitrificans
C;Superfamily: precorrin-3 methylase

Query Match 19.0%; Score 61.5; DB 2; Length 253;
Best Local Similarity 32.3%; Pred. No. 16;
Matches 21; Conservative 14; Mismatches 23; Indels 7; Gaps 3;

QY 4 IIGTPTWEOQAAAEAVQARDSDVDIRVARVIEODMAVDSAGKITRYIKLE---VSFKM 60
Db 51 IVDTAPMSLDIEAEYVVAEAGLD---VARLHSGDSVWVAEAEQIR-RLEKHGIATM 106

QY 61 RPAQP 65
Db 107 TPGVP 111

RESULT 13
EWL18
EI protein - human papillomavirus type 18
C;Species: human papillomavirus type 18
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C;Accession: C26165; C26251
R. Seedorf, K.; Oltersdorf, T.; Kraemer, G.; Roewekamp, W.
EMBO J. 6, 139-144, 1987
A;Title: Identification of early proteins of the human papilloma viruses type 16 (HPV 16
A;Reference number: A91068; MUID:87218459; PMID:3034571
A;Accession: C26165
A;Molecule type: DNA
A;Residues: 1-274 <SES>
A;Cross-references: GB:X04773; NID:G60876; PIDN:CAA28468.1; PID:G60879
R. Cole, S.T.; Panos, O.
J. Mol. Biol. 193, 599-608, 1987
A;Title: Nucleotide sequence and comparative analysis of the human papillomavirus type 1
A;Reference number: A92937; MUID:87283982; PMID:3039146
A;Accession: C26251
A;Molecule type: DNA
A;Residues: 1-657 <COL>

A;Cross-references: GB:X05015; NID:G60975; PIDN:CAA28666.1; PID:G60978
C;Superfamily: papillomavirus E1 protein
C;Keywords: early protein

Query Match 19.0%; Score 61.5; DB 1; Length 657;
Best Local Similarity 29.2%; Pred. No. 44;
Matches 19; Conservative 14; Mismatches 29; Indels 3; Gaps 1;

QY 1 VIDIGTPTWEOQAAAEAVQARDSDVDIRVARVIEODMAVDSAGKITRYIKLEVS 57
Db 47 WDFIDTQTTFEQAELETAQALFHAQEVHNDQVHLVKRFAGGSTENSPIGSERLEVD 106

QY 58 FKMRP 62
Db 107 TELSP 111

RESULT 14
S08243
beta-glucosidase (EC 3.2.1.21) - Ruminococcus albus
C;Species: Ruminococcus albus
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Oct-1999
C;Accession: S08243
R. Ohmura, K.; Takano, M.; Shimizu, S.
Nucleic Acids Res. 18, 671, 1990
A;Title: DNA sequence of a beta-glucosidase from Ruminococcus albus.
A;Reference number: S08243; MUID:90175009; PMID:2106673
A;Accession: S08243
A;Molecule type: DNA
A;Residues: 1-947 <OHM>
A;Cross-references: EMBL:X15415; NID:G45967; PIDN:CAA33461.1; PID:G45968
C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 19.0%; Score 61.5; DB 2; Length 947;
Best Local Similarity 33.3%; Pred. No. 66;
Matches 18; Conservative 11; Mismatches 22; Indels 3; Gaps 2;

QY 13 EQAAAEAVQARDSDVDIRVARV--IEQDMAVDSAGKITRYIKLEVSFQMRQAQ 64
Db 774 EDEAVEVINKPAETVDDGEGDRVFLDGLTDIMSGVKTER-NLDYSFTVDVAQ 826

RESULT 15
SYECCP
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) large chain [validated]
N;Alternate names: carbamoyl-phosphate synthetase (glutamine-hydrolyzing) ammonia chain
C;Species: Escherichia coli
C;Date: 19-Feb-1984 #sequence_revision 31-Dec-1991 #text_change 01-Mar-2002
C;Accession: A01198; S40556; A64724
R. Nyunoya, H.; Lusty, C.J.
Proc. Natl. Acad. Sci. U.S.A. 80, 4629-4633, 1983
A;Title: The carb gene of Escherichia coli: a duplicated gene coding for the large subunit
A;Reference number: A01198; MUID:83273669; PMID:8308632
A;Accession: A01198
A;Molecule type: DNA
A;Residues: 1-1073 <NYU>
A;Cross-references: GB:J01597; NID:G145461; PIDN:AAA23539.1; PID:G145464
R. Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobuc
Submitted to the EMBL Data Library, December 1992
A;Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2.
A;Reference number: S40531
A;Accession: S40556
A;Molecule type: DNA
A;Residues: 1-1073 <YUR>
A;Cross-references: EMBL:D10483; NID:G216434; PIDN:BA01311.1; PID:G216460
R. Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A64724
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

This page blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 0.778231 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-6

Perfect score: 324

Sequence: 1 VIDIIIGTSPTEQAAAEV.....KITRIKLEVSFQRPAPQR 56

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	65.5	20.2	1	CARB_ECO57
2	63.5	19.6	1	ID12_RICCN
3	63.5	19.6	1	ID12_RICPR
4	63	19.4	1	HUTU_CABEL
5	61.5	19.0	253	1 COB_M_PSRDE
6	61.5	19.0	657	1 VBI_HFV18
7	61.5	19.0	947	1 BGLS_RUMAL
8	61.5	19.0	1072	1 CARB_ECOLI
9	61	18.8	191	1 RS7_BRAOL
10	60.5	18.7	351	1 ID12_RHILQ
11	60.5	18.7	1074	1 CARB_SALTI
12	60.5	18.7	1074	1 CARB_SALTY
13	60	18.5	321	1 KGF_P_PASNU
14	59.5	18.4	458	1 DESM_XENLA
15	59.5	18.4	959	1 SYL_AERPE
16	59.5	18.4	988	1 E4L2_MOUSE
17	59	18.2	505	1 SCRY_SALTY
18	58.5	18.1	361	1 COBT_MYCTU
19	58.5	18.1	1005	1 E4L2_HUMAN
20	58	17.9	974	1 TP4_MOUSE
21	58	17.9	977	1 TRP4_HUMAN
22	58	17.9	977	1 TRP4_RAT
23	58	17.9	981	1 TRP4_BOVIN
24	58	17.9	1577	1 MYSH_ACACA
25	57.5	17.7	286	1 PANB_RHILQ
26	57.5	17.7	501	1 GRIS_HUMAN
27	57.5	17.7	1078	1 CARB_BUCAI
28	57	17.6	1184	1 DP3A_MYCTU
29	56.5	17.4	355	1 VPI_HAEIN
30	56.5	17.4	1332	1 YSVI_CABEL
31	56	17.3	473	1 SVC_METMA
32	56	17.3	515	1 LBUI_HAEIN
33	56	17.3	754	1 ASFH_BOVIN

ALIGNMENTS

```

RESULT 1
CARB_ECO57
ID CARB_ECO57 STANDARD; PRT; 1072 AA.
AC Q9XA38;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE Phosphate synthetase ammonia chain).
GN CARB OR Z0038 OR ECS0036 OR SF0030.
OS Escherichia coli O157:H7, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334, 623;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Rose D.J., Plunkett G., Burtland V., Mau B., Glasner J.D.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.F., Potamowitz K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Liida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Khara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
RN [3]
SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
CC -I- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -I- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -I- PATHWAY: Arginine biosynthesis.
CC -I- PATHWAY: Pyrimidine biosynthesis; first step.
CC -I- SUBUNIT: Composed of two chains; the small (or glutamine) chain

```

```

34 55.5 17.1 149 1 RBFA_COREF Q8fpa8 corynebacte
35 55.5 17.1 251 1 LIPB_COREF Q8fnp5 corynebacte
36 55.5 17.1 253 1 NIKD_ECOLI P33593 escherichia
37 55.5 17.1 464 1 AAKI_HUMAN Q9ugi9 homo sapien
38 55.5 17.1 651 1 BGLR_HUMAN P08236 homo sapien
39 55.5 17.1 658 1 VGL8_BPT4 P13332 bacterioph
40 55.5 17.1 699 1 AT6B_MOUSE O35451 mus musculu
41 55 17.0 296 1 THT2_MYCLE P46700 mycobacteri
42 55 17.0 395 1 PELO_DROME P48612 drosophila
43 55 17.0 402 1 GRSB_BACSU P71017 bacillus su
44 55 17.0 532 1 CRTI_APHSP P21134 aphanocapsa
45 55 17.0 689 1 UVRB_HAUNI Q9hmt9 halobacteri

```

CC promotes the hydrolysis of glutamine to ammonia, which is used by
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate.
 CC Tetramer of heterodimers (alpha,beta)4 (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; A005180; AAC54335.1; -
 CC EMBL; A002550; BAB33459.1; -
 CC EMBL; A015040; AAN41696.1; -
 CC PIR; C85484; C85484.
 CC PIR; D90633; D90633.
 CC HAMAP; MF_01210; -; 1.
 CC InterPro; IPR006275; Cpsase_L_glu.
 CC InterPro; IPR005483; Cpsase_L.
 CC InterPro; IPR005479; Cpsase_L_D2.
 CC InterPro; IPR005480; Cpsase_L_D3.
 CC InterPro; IPR005481; Cpsase_L_N.
 CC InterPro; IPR004382; MGS_L1.
 CC Pfam; PF00289; Cpsase_L_chain; 2.
 CC Pfam; PF02786; Cpsase_L_D2; 2.
 CC Pfam; PF02787; Cpsase_L_D3; 2.
 CC Pfam; PF02142; MGS; 1.
 CC PRINTS; PR00098; CPSASE.
 CC TIGRfam; TIGR01369; CpsaseII_lrg; 1.
 CC PROSITE; PS00866; CPSASE_1; 2.
 CC PROSITE; PS00867; CPSASE_2; 2.
 CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; Complete proteome.
 FT INIT MET 0
 FT DOMAIN 1 402
 FT DOMAIN 403 552
 FT DOMAIN 553 935
 FT DOMAIN 936 1072
 FT REPEAT 1 552
 FT NP_BIND 152 209
 FT NP_BIND 302 353
 FT METAL 284 284
 FT METAL 288 288
 FT METAL 300 300
 FT METAL 828 828
 FT METAL 840 840
 SQ SEQUENCE 1072 AA; 117724 MW; D98867964E058E37 CRC64;
 Query Match 20.2%; Score 65.5; DB 1; Length 1072;
 Best Local Similarity 30.6%; Pred. No. 20;
 Matches 19; Conservative 17; Mismatches 19; Indels 7; Gaps 2;
 QY 2 IDIIGTSPTSEQAAAEVQARDSDIVR-----ARVIEQDMADVDSAGKITRYIKLEV 56
 Db 658 VPIVIGTSPDAIR--AEDRRFQHAVDRLKQKPANATVTAIEAVKAKKIGYPLVVRP 715
 QY 57 SF 58
 Db 716 SY 717
 RESULT 2
 IDI2_RICCN STANDARD; PRT; 342 AA.
 AC Q92HW7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2) (IPP isomerase)
 DE (Isopentenyl pyrophosphate isomerase).

FN1 OR RC0744.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsieae; Rickettsia.
 OX NCBI_TaxID=781;
 RN SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 EX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Resasco-Audiffren P., Fournier P.-E., Barbe V.,
 RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Rault D.;
 RA "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 CC -!- FUNCTION: Catalyzes the 1,3-allylic rearrangement of the
 CC homoallylic substrate isopentenyl (IPP) to its allylic isomer,
 CC dimethylallyl diphosphate (DMAPP) (By similarity).
 CC -!- CATALYTIC ACTIVITY: Isopentenyl diphosphate = dimethylallyl
 CC diphosphate.
 CC -!- COFACTOR: FMN and NADPH (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the IPP isomerase type 2 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE008632; AAL03282.1; -
 CC PIR; H97792; H97792.
 CC HAMAP; MF_00354; -; 1.
 CC InterPro; IPR003009; FMN enzyme.
 KW Isomerase; Isoprene biosynthesis; Flavoprotein; FMN; NADP;
 KW Complete proteome.
 SQ SEQUENCE 342 AA; 37336 MW; D346003CCD99B7A3 CRC64;
 Query Match 19.6%; Score 63.5; DB 1; Length 342;
 Best Local Similarity 25.8%; Pred. No. 11;
 Matches 16; Conservative 13; Mismatches 14; Indels 19; Gaps 2;
 QY 1 VDIIGTSPTSEQAAAEVQARDSDIVR-----VDIRVARVIEQDMADVDS 44
 Db 215 VLDIAGSGGTSQVEA---YRATNSLQNRRIASSFTNWGIPITLDSLKMVREVSQDIPIT 271
 QY 45 AG 46
 Db 272 SG 273
 RESULT 3
 IDI2_RICPR STANDARD; PRT; 342 AA.
 AC Q9ZD30;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2) (IPP isomerase)
 DE (Isopentenyl pyrophosphate isomerase).
 GN FN1 OR RP452.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsieae; Rickettsia.
 OX NCBI_TaxID=782;
 RN SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria.";
CC -!- FUNCTION: CATALYZES THE 1,3-ALLYLIC REARRANGEMENT OF THE
CC HOMOALLYLIC SUBSTRATE ISOPENTENYL (IPP) TO ITS ALLYLIC ISOMER,
CC DIMETHYLLALLYL DIPHOSPHATE (DMAPP) (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: isopentenyl diphosphate = dimethylallyl
CC diphosphate.
CC -!- COFACTOR: FMN AND NADPH (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the IPP isomerase type 2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ235271; CAAL4909.1; -.
CC PIR: C71704; C71704.
CC HAVAP; MF 00354; -; 1.
CC InterPro: IPR003009; FMN enzyme.
CC KX Isomerase; Isoprene biosynthesis; Flavoprotein; FMN; NADP;
CC Complete proteome.
CC KW Complete proteome.
CC SQ SEQUENCE 342 AA; 37363 MW; FE452A843CAlB950 CRC64;

Query Match 19.6%; Score 63.5; DB 1; Length 342;
Best Local Similarity 25.8%; Pred. No. 11;
Matches 16; Conservative 14; Mismatches 13; Indels 19; Gaps 2;

QY 1 VIDIGTSPTSWEQAAEAQVARDSDS-----VDDIRVARVIRQDVAVD 44
DB 215 VLDIAGSGTWSQVEA---YRAKSNQNRFIASFWGTTIDSLKMLQEI SKDTITIA 271
QY 45 AG 46
DB 272 SG 273

RESULT 4
HUTU_CAEEL
ID HUTU_CAEEL STANDARD; PRT; 731 AA.
AC Q9NAE2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable urocanate hydratase (EC 4.2.1.49) (Urocanase)
DE (Imidazolepropionate hydrolase).
GN Y51H4A.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Sulston J.E.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: 3-(5-oxo-4,5-dihydro-3-H-imidazol-4-
CC Y)propanoate = urocanate + H(2)O.
CC -!- COFACTOR: NAD (By similarity).
CC -!- PATHWAY: Histidine degradation; second step.
CC -!- SIMILARITY: BELONGS TO THE UROCANASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

CC EMBL: AL132952; CAB61139.1; -.
DR WormPep; Y51H4A.7; CE22333.
DR InterPro: IPR000193; Urocanase.
DR Pfam: PF01175; Urocanase; 1.
DR ProDom: PD025423; Urocanase; 1.
DR PROSITE: PS01233; UROCANASE; 1.
KW Hypothetical protein; Histidine metabolism; Lyase; NAD.
SQ SEQUENCE 731 AA; 81523 MW; 891EF61C9A604E6 CRC64;

Query Match 19.4%; Score 63; DB 1; Length 731;
Best Local Similarity 41.7%; Pred. No. 25;
Matches 15; Conservative 4; Mismatches 15; Indels 2; Gaps 1;

QY 6 GTSPTSW--EQAAEAQVARDSDVDDIRVARVIRQD 39
DB 687 GVTRESWSGNAKAQAIQRAEKQVDGLRVTLPVEAD 722

RESULT 5
COBM_PSEDE
ID COBM_PSEDE STANDARD; PRT; 253 AA.
AC P21922;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Precorrin-4 C11-methyltransferase (EC 2.1.1.133) (Precorrin-3
DE methylase).
DE methyllase).
GN COBM.
OS Pseudomonas denitrificans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=43306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC510;
RA Crouzet J., Thibaut D., Cauchois L., Rigault S., Rouyez M.-C.,
RA Blanche F., Thibaut D., Debussche L.;
RT "Genetic and sequence analysis of an 8.7-kilobase Pseudomonas
RT denitrificans fragment carrying eight genes involved in
RT transformation of precorrin-2 to cobyrinic acid.";
RL J. Bacteriol. 172:5980-5990(1990).
RN [2]
RP CHARACTERIZATION.
RC MEDLINE=94042919; PubMed=8226690;
RA Debussche L., Thibaut D., Cameron B., Crouzet J., Blanche F.J.;
RT "Biosynthesis of the corrin macrocycle of coenzyme B12 in Pseudomonas
RT denitrificans.";
RL J. Bacteriol. 175:7430-7440(1993).
CC -!- FUNCTION: METHYLTRANSFERASE THAT CATALYZES THE METHYLATION OF
CC C-11 IN PRECORRIN-4 TO FORM PRECORRIN-5.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-4 = S-
CC adenosyl-L-homocysteine + precorrin 5.
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC AND CBIL/COBI.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M59301; AAA25801.1; -.
DR InterPro: IPR006362; COB.
DR InterPro: IPR000878; Cor/por Metransf.
DR InterPro: IPR003043; Uropor Metransf.
DR Pfam: PF00590; TP methylase; 1.
DR TIGRFAMs: TIGR01465; coBM cbif; 1.
DR PROSITE: PSC0839; SUMT_1; 1.

RESULT 8

ID CARB_ECOLI STANDARD; PERT; 1072 AA.
 AC P00968;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
 GN CARB OR PYRA OR B0033.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RC SEQUENCE FROM N.A., AND SEQUENCE OF 1-1.
 RC STRAIN=K12;
 RX MEDLINE=83273669; PubMed=6308632;
 RA Nyunoya H., Lusty C.J.;
 RT "The carb gene of Escherichia coli: a duplicated gene coding for the large subunit of carbamoyl-phosphate synthetase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4629-4633(1983).
 RN [2]
 RC SEQUENCE FROM N.A.
 RP MEDLINE=94248073; PubMed=6377309;
 RX Bouvier J., Patte J.-C., Stragier P.;
 RA "Multiple regulatory signals in the control region of the Escherichia coli carbA operon.";
 RT Proc. Natl. Acad. Sci. U.S.A. 81:4139-4143(1984).
 RN [3]
 RC SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=92334977; PubMed=1630901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N., Isono K., Mizobuchi K., Nakata A.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of the 0-2.4 min region.";
 RL Nucleic Acids Res. 20:3305-3308(1992).
 RN [4]
 RC SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [5]
 RC SEQUENCE OF 1-6 FROM N.A.
 RP MEDLINE=84248072; PubMed=6330744;
 RX Piette J., Nyunoya H., Lusty C.J., Cunin R., Weyens G., Crabeel M., Charlier D.R.M., Glansdorff N., Piérard A.;
 RA "DNA sequence of the carb gene and the control region of carbA: tandem promoters, respectively controlled by arginine and the pyrimidines, regulate the synthesis of carbamoyl-phosphate synthetase in Escherichia coli K-12.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:4134-4138(1984).
 RN [6]
 RC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RP MEDLINE=97317071; PubMed=9174345;
 RA Thoden J.B., Holden H.M., Wesenberg G., Raushel F.M., Rayment I.;
 RT "Structure of carbamoyl phosphate synthetase: a journey of 96 A from substrate to product.";
 RL Biochemistry 36:6305-6316(1997).
 RN [7]
 RC X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RP MEDLINE=98301363; PubMed=9636022;
 RA Thoden J.B., Miran S.G., Phillips J.C., Howard A.J., Raushel F.M., Holden H.M.;
 RT "Carbamoyl phosphate synthetase: caught in the act of glutamine

hydrolysis.";
 RL Biochemistry 37:8825-8831(1998).
 RN [8]
 RC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RP MEDLINE=99190825; PubMed=10089390;
 RA Thoden J.B., Raushel F.M., Benning M.M., Rayment I., Holden H.M.;
 RT "The structure of carbamoyl phosphate synthetase determined to 2.1-A resolution.";
 RL Acta Crystallogr. D 55:8-24(1999).
 RN [9]
 RC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RP MEDLINE=99152507; PubMed=10029528;
 RA Thoden J.B., Wesenberg G., Raushel F.M., Holden H.M.;
 RT "Carbamoyl phosphate synthetase: closure of the B-domain as a result of nucleotide binding.";
 RL Biochemistry 38:2347-2357(1999).
 RN [10]
 RC X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RP MEDLINE=20056035; PubMed=10587438;
 RA Thoden J.B., Huang X., Raushel F.M., Holden H.M.;
 RT "The small subunit of carbamoyl phosphate synthetase: snapshots along the reaction pathway.";
 RL Biochemistry 38:16158-16166(1999).
 RN [11]
 RC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RP MEDLINE=99357782; PubMed=10428826;
 RA Thoden J.B., Raushel F.M., Wesenberg G., Holden H.M.;
 RT "The binding of inosine monophosphate to Escherichia coli carbamoyl phosphate synthetase.";
 RL J. Biol. Chem. 274:22502-22507(1999).
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- COFACTOR: Binds 3 manganese ions per subunit.
 CC -!- PATHWAY: Arginine biosynthesis.
 CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate.
 CC Tetramer of heterodimers (alpha,beta)4.
 CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; V01500; CA24744.1; -.
 DR EMBL; D10483; BAB96602.1; -.
 DR EMBL; AE000113; AAC73144.1; -.
 DR PIR; A01198; SYECCP.
 DR PDB; 1JDB; 17-JUN-98.
 DR PDB; 1A9X; 21-OCT-98.
 DR PDB; 1BX8; 20-APR-99.
 DR PDB; 1CS8; 26-JUL-99.
 DR PDB; 1C30; 10-DEC-99.
 DR PDB; 1CS0; 10-DEC-99.
 DR PDB; 1KEE; 21-DEC-01.
 DR PDB; 1M6V; 13-NOV-02.
 DR ECODBASE; E1330; 6TH EDITION.
 DR Ecogene; EGI0135; carb.
 DR HAMAP; MF_01210; -; 1.
 DR InterPro; IPR006275; CarA_L_glu.
 DR InterPro; IPR005483; CPase_L.
 DR InterPro; IPR005479; CPase_L_D2.
 DR InterPro; IPR005480; CPase_L_D3.
 DR InterPro; IPR005481; CPase_L_N.
 DR InterPro; IPR004362; MGS_like.
 DR Pfam; PF0289; CPase_L_chain; 2.
 DR Pfam; PF02786; CPase_L_D2; 2.

DR Pfam; PF02787; CPSase_L_D3; 1.
 DR Pfam; PF02142; MGS; 1.
 DR PRINTS; PRO0098; CPSASE.
 DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
 DR PROSITE; PS00866; CPSASE_1; 2.
 DR PROSITE; PS00867; CPSASE_2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; 3D-structure; Complete proteome.
 FT INIT MET 0
 FT DOMAIN 1 402 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 403 552 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 553 935 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 936 1072 ALLOSTERIC DOMAIN.
 FT REPEAT 1 552
 FT REPEAT 553 1072
 FT NP_BIND 152 209
 FT NP_BIND 302 353
 FT NP_BIND 354 384
 FT METAL 298 298
 FT METAL 299 300
 FT METAL 300 300
 FT METAL 828 828
 FT METAL 840 840
 FT STRAND 8 12
 FT STRAND 18 18
 FT TURN 19 20
 FT STRAND 21 21
 FT TURN 23 23
 FT TURN 24 38
 FT TURN 39 40
 FT STRAND 42 46
 FT TURN 50 51
 FT TURN 51 51
 FT HELIX 53 55
 FT HELIX 57 59
 FT STRAND 62 64
 FT HELIX 70 80
 FT STRAND 84 86
 FT STRAND 91 103
 FT TURN 104 105
 FT HELIX 106 109
 FT TURN 110 111
 FT STRAND 113 114
 FT HELIX 119 126
 FT HELIX 128 137
 FT TURN 138 139
 FT TURN 145 148
 FT HELIX 151 161
 FT STRAND 165 169
 FT TURN 170 171
 FT TURN 174 177
 FT STRAND 179 181
 FT HELIX 184 197
 FT TURN 199 200
 FT STRAND 203 207
 FT TURN 210 211
 FT STRAND 213 221
 FT TURN 223 224
 FT STRAND 227 235
 FT TURN 239 240

Query Match 19.0%; Score 61.5; DB 1; Length 1072;
 Best Local Similarity 29.0%; Pred. NO. 53;
 Matches 18; Conservative 18; Mismatches 19; Indels 7; Gaps 2;
 QY 2 IDIIGTSPTSWEQAAAEAVQARDSDVDIRV-----ARVIEDQNAVDASAGKITRYIKLEVSF 56
 Db 658 VPIVIGTSPIADR--AEDREFQHAVERLKLKQPANATVTAIENAVEKAKEIGPLVLRP 715

QY 57 SF 58
 Db 716 SY 717

RESULT 9

RS7_BRAOL
 ID RS7_BRAOL STANDARD; PRT; 191 AA.
 AC Q9XK45;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 40S ribosomal protein S7.
 GN RPS7.
 OS Brassica oleracea (Cauliflower).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 NCBI_TaxID=3712;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gossen K.K., Katavic V., Taylor D.C.;
 RT "Nucleotide sequence of a cDNA from microspore derived embryos of
 RT Brassica oleracea is homologous to 40S ribosomal protein S7.";
 RL (in) Plant Gene Register PGR99-096.
 CC -1- SIMILARITY: BELONGS TO THE S7E FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF144752; AAD44761.1; -
 DR InterPro; IPR000554; Ribosomal_S7E.
 DR Pfam; PF01251; Ribosomal_S7e; 1.
 DR ProDom; PD006276; Ribosomal_S7e; 1.
 DR PROSITE; PS00948; RIBOSOMAL_S7E; FALSE_NEG.
 KW Ribosomal protein.
 SQ SEQUENCE 191 AA; 22154 MW; 175A2D794210077C CRC64;
 Query Match 18.8%; Score 61; DB 1; Length 191;
 Best Local Similarity 23.1%; Pred. NO. 11;
 Matches 15; Conservative 15; Mismatches 23; Indels 12; Gaps 2;
 QY 6 GTSPTSWEQAAAEAV-----QARDSDVDIRVARVIBQDVAVDASAGKITRYIKLEVSF 58
 Db 13 GVAPTEFEERVAQAPDLENTNQELKSLDKLYINQAVSMIDAGNEKAVIY-----VVF 67
 QY 59 KMRPA 63
 Db 68 RLAKA 72
 RESULT 10
 IDI2_RHILO
 ID IDI2_RHILO STANDARD; PRT; 351 AA.
 AC Q899L5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2) (IPP isomerase)
 DE [isopentenyl pyrophosphate isomerase].
 GN FNI OR MNR6371.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NAF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;

```

J. Bacteriol. 185:2330-2337(2003)).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation. It
CC is the property of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR ENBL; AL627265; CAD01221.1; -.
DR ENBL; AE016834; AAO67801.1; -.
DR HAMAP; MF_01210; -.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005493; CPhase_L.
DR InterPro; IPR005479; CPhase_L_D2.
DR InterPro; IPR005480; CPhase_L_D3.
DR InterPro; IPR005481; CPhase_L_N.
DR InterPro; IPR004362; MGS-like.
DR Pfam; PF00289; CPhase_L_chain; 2.
DR Pfam; PF02786; CPhase_L_D2; 2.
DR Pfam; PF02787; CPhase_L_D3; 1.
DR Pfam; PF02142; MGS_1; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRams; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00857; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT INIT_MET 0
FT DOMAIN 1 402 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 552 OLIGOMERIZATION DOMAIN.
FT DOMAIN 553 935 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 936 1074 ALLOSTERIC DOMAIN.
FT REPEAT 1 552
FT REPEAT 553 1074 ATP (POTENTIAL).
FT NP_BIND 152 209 ATP (POTENTIAL).
FT NP_BIND 302 353 MANGANESE 1 (BY SIMILARITY).
FT METAL 284 288 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 3 (BY SIMILARITY).
FT METAL 828 828 MANGANESE 3 (BY SIMILARITY).
FT METAL 840 840 MANGANESE 3 (BY SIMILARITY).
FT SEQUENCE 1074 AA, 118009 MW, F69DE79502C4CD15 CRC64;
SQ
Query Match 18.7%; Score 60.5; DB 1; Length 1074;
Best Local Similarity 29.0%; Pred.No.68;
Matches 18; Conservative 17; Mismatches 20; Indels 7; Gaps 2;
QY 2 IDIVGSPTSWEQAAAEAVQRARSDVDDIRV-----ARVIEQDMADVDSAGKITRYIKLEV 56
Db 658 VPIGTSPDAIDR--AEDRRERFQAVDRLLKPKPANTVTVAIQAVKAKEIGYPLVVRP 715
QY 57 SF 58
Db 716 SY 717
-----
RESULT 12
CARE SALTY
ID -CARE_SALTY STANDARD; PRT; 1074 AA.
AC p14836; p96067;
OT 01-APP-1990 (Rel. 14. Created)

```

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain)
 GN CARB OR STM0067
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RA Lu C.D., Walther D.A., Abdelal A.T.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
 RT Nature 413:852-856 (2001).
 RN [3]
 RP SEQUENCE OF 1-7 FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=88329100; PubMed=2843375;
 RA Kilstrup M., Lu C.D., Abdelal A., Neuhaud J.;
 RT "Nucleotide sequence of the carA gene and regulation of the carAB operon in Salmonella typhimurium.";
 RL Eur. J. Biochem. 176:421-429 (1988).
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -!- PATHWAY: Arginine biosynthesis.
 CC -!- PATHWAY: Pyrimidine biosynthesis;
 CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate.
 CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; U81260; AAB339256.1; -
 DR EMBL; AE008696; AAL19031.1; -
 DR EMBL; M36540; AAB27033.1; -
 DR HSSP; P00968; IABX.
 DR StyGene; SG10033; carB.
 DR HAMAP; MF_01210; -; 1.
 DR InterPro; IPR006275; CarA_L_glu.
 DR InterPro; IPR005483; CPase_L.
 DR InterPro; IPR005479; CPase_L_D2.
 DR InterPro; IPR005480; CPase_L_D3.
 DR InterPro; IPR005481; CPase_L_N.
 DR InterPro; IPR004362; MGS_like.
 DR Pfam; PF00289; CPase_L_chain; 2.
 DR Pfam; PF02786; CPase_L_D2; 2.
 DR Pfam; PF02787; CPase_L_D2; 2.
 DR Pfam; PF02142; MGS; 1.
 DR PRINTS; PR00098; CPASE.
 DR TIGRFAMs; TIGR01369; CPaseII_lrg; 1.
 DR PROSITE; PS00866; CPASE_1; 2.
 DR PROSITE; PS00867; CPASE_2; 2.
 DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW

KW ATP-binding; Manganese; Complete proteome.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT DOMAIN 1 402 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 403 552 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 553 935 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 936 1074 ALLOSTERIC DOMAIN.
 FT REPEAT 1 552
 FT REPEAT 553 1074 ATP (POTENTIAL).
 FT NP_BIND 152 209 ATP (POTENTIAL).
 FT NP_BIND 302 353 MANGANESE 1 (BY SIMILARITY).
 FT METAL 284 284 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 298 298 MANGANESE 2 (BY SIMILARITY).
 FT METAL 300 300 MANGANESE 3 (BY SIMILARITY).
 FT METAL 828 828 MANGANESE 3 (BY SIMILARITY).
 FT METAL 840 840 MANGANESE 3 (BY SIMILARITY).
 FT CONFLICT 122 122 I -> G (IN REF. 1).
 SQ SEQUENCE 1074 AA; 118007 MW; 19243FECE2C8D5C7 CRC64;
 Query Match 18.7%; Score 60.5; DB 1; Length 1074;
 Best Local Similarity 29.0%; Pred. No. 68;
 Matches 18; Conservative 17; Mismatches 20; Indels 7; Gaps 2;
 QY 2 IDIIGTSPSWQAARAEVQRADSVDDIRV-----ARVIEQDAVDSAGKITRYIKLEV 56
 DB 658 VPIGTSPDAIDR--AEDRERFQAVDRILKQPNATVTAIEQAVEKAKEIGYPLVVRP 715
 QY 57 SF 58
 DB 716 SY 717
 RESULT 13
 K6PF PASMU STANDARD; PRT; 321 AA.
 ID K6PF PASMU STANDARD; PRT; 321 AA.
 AC QSCP2; 2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)
 GN PFKA OR PM0069
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70.
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Faustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
 CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate.
 CC -!- PATHWAY: Key control step of glycolysis.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the phosphofructokinase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; AE006042; AAK02153.1; -
 DR HSSP; P06998; 2PFK.
 DR HAMAP; MF_00339; -; 1.
 DR InterPro; IPR000023; Ppfckinase.
 DR Pfam; PF00365; PFK; 1.
 DR PRINTS; PR00476; PPFCKTKINASE.
 DR ProDom; PD000707; Ppfckinase; 1.

[illegible]

Db 833 VLSLIGGADTLVTVVAAEWKYRAVEAVRRRERGASMKKEALREAFKVEGVDKREARLVQ 892

QY 38 Q 38

Db 893 Q 893

Search completed: November 21, 2003, 16:04:17
Job time : 1.77823 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 , Search time 3.24762 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-6
Perfect score: 324
Sequence: 1 VIDIIIGTSPISWEQAAAEAV.....KITVRIKLEVSFKMRPAQPR 66

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:**

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	324	100.0	71	16	053714
2	319	98.5	116	2	007267
3	265	81.8	65	2	007265
4	190	58.6	110	2	006124
5	140	43.2	96	2	006125
6	101	31.2	77	17	Q9HPW4
7	92	28.4	110	16	Q8RAA6
8	76	23.5	70	16	Q8KAD4
9	69	21.3	71	16	Q9I794
10	66	20.4	69	16	Q92X94
11	65	20.1	83	16	Q8XQ96
12	64.5	19.9	501	6	Q9WNL1
13	64.5	19.9	1030	5	Q962B3
14	64	19.8	96	16	Q8Z0H1
15	64	19.8	206	15	Q75741
16	64	19.8	523	5	Q9NE33

17	64	19.8	1180	3	Q8J0Z4
18	64	19.8	1200	3	Q8J122
19	63	19.4	85	16	Q9RCZ5
20	63	19.4	155	12	Q39829
21	63	19.4	526	3	Q8X0K0
22	62.5	19.3	436	2	Q8RPT8
23	62.5	19.3	456	12	Q8UL22
24	62	19.1	282	2	O52068
25	62	19.1	282	16	Q9YDMS
26	62	19.1	282	16	Q8FUS2
27	62	19.1	1090	2	Q8RPF1
28	62	19.1	1094	5	Q23915
29	61.5	19.0	252	16	Q92LV3
30	61.5	19.0	274	12	Q9AE26
31	61.5	19.0	335	5	Q8T3W1
32	61.5	19.0	339	5	Q8INJ1
33	61.5	19.0	477	5	Q8INJ2
34	61.5	19.0	482	16	Q8FOL4
35	61.5	19.0	495	5	Q9VGD7
36	61.5	19.0	1073	16	Q8FLB0
37	61.5	19.0	1477	5	Q9VB52
38	61.5	19.0	1477	5	O76931
39	61.5	19.0	94	17	Q9HRH0
40	61	18.8	397	2	Q9AG69
41	61	18.8	2626	11	Q8ZLN3
42	60.5	18.7	373	2	Q8KJK8
43	60.5	18.7	481	2	Q9X754
44	60.5	18.7	562	9	Q9G064
45	60.5	18.7			

ALIGNMENTS

RESULT 1

OS3714 ID OS3714 PRELIMINARY; PRT; 71 AA.
AC OS3714;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Possible protein TRANSPORT protein.
GN SEC OR RV0379 OR MTU036.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
DR EMBL; AL021931; CAA17385.1;
DR TuberculList; RV0379;
KW Complete proteome.
SQ SEQUENCE 71 AA; 7966 MW; 49526FDBCA9C3826 CRC64;

Query Match 100.0%; Score 324; DB 16; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIIGTSPISWEQAAAEAVQSRQSDVDIRVARVIEDQMAVDASAGKITVRIKLEVSFKMR 60
|||||

```
Db 6 VIDIIGTSPTEWQAAAEAVQARSDVDDIRVARVIEQDVAQDSAGKITRYIKLEVSFKM 65
QY 61 RPAQPR 66
Db 66 RPAQPR 71

RESULT 2
O07267 PRELIMINARY; PRT; 116 AA.
AC O07267;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE Hypothetical 12.7 kDa protein (Fragment).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TB patient sputum #NTI64719;
RA Rao A.R., Vijaya S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U85467; AAB58554.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 116 AA; 12673 MW; DF56D2FF0B479896 CRC64;

Query Match 98.5%; Score 319; DB 2; Length 116;
Best Local Similarity 98.5%; Pred. No. 1e-28;
Matches 65; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEWQAAAEAVQARSDVDDIRVARVIEQDVAQDSAGKITRYIKLEVSFKM 60
Db 51 VIDIIGTSPTEWQAAAEAVQARSDVDDIRVARVIEQDVAQDSAGKITRYIKLEVSFKM 110
QY 61 RPAQPR 66
Db 111 RPAQPR 116

RESULT 3
O07265 PRELIMINARY; PRT; 65 AA.
AC O07265;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE Hypothetical 7.2 kDa protein (Fragment).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TB patient sputum #NTI64719;
RA Rao A.R., Vijaya S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U85465; AAB58552.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 65 AA; 7240 MW; 9339C19E2306FA1B CRC64;

Query Match 81.8%; Score 265; DB 2; Length 65;
Best Local Similarity 85.9%; Pred. No. 7.3e-23;
Matches 55; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 IDIIGTSPTEWQAAAEAVQARSDVDDIRVARVIEQDVAQDSAGKITRYIKLEVSFKM 61
Db 1 IDIIGTSPTEWQAAAEAVQARSDVDDIRVARVIEQDVAQDSAGKITRYIKLEVSFKM 60
QY 62 PAQP 65
```

```
Db 61 PAQP 64

RESULT 4
O06124 PRELIMINARY; PRT; 110 AA.
AC O06124;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE Hypothetical 11.9 kDa protein (Fragment).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTI 64719;
RA Sachidanandam V., Rama Rao A., Hegde S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001301; AAB54027.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 110
SQ SEQUENCE 110 AA; 11903 MW; A21AB7D99A7B4F54 CRC64;

Query Match 58.8%; Score 190; DB 2; Length 110;
Best Local Similarity 95.2%; Pred. No. 4.8e-14;
Matches 40; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEWQAAAEAVQARSDVDDIRVARVIEQDVAQDSAGKITRYIKLEVSFKM 42
Db 53 VIDIIGTSPTEWQAAAEAVQARSDVDDIRVARVIEQDVAQDSAGKITRYIKLEVSFKM 94

RESULT 5
O06125 PRELIMINARY; PRT; 96 AA.
AC O06125;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE Hypothetical 10.3 kDa protein (Fragment).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTI 64719;
RA Sachidanandam V., Rama Rao A., Hegde S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001302; AAB54028.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 96
SQ SEQUENCE 96 AA; 10298 MW; 2E5B14CB1D1E445C CRC64;

Query Match 43.2%; Score 140; DB 2; Length 96;
Best Local Similarity 86.1%; Pred. No. 2e-08;
Matches 31; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEWQAAAEAVQARSDVDDIRVARVIEQDVAQDSAGKITRYIKLEVSFKM 36
Db 61 VIDIIGTSPTEWQAAAEAVQARSDVDDIRVARVIEQDVAQDSAGKITRYIKLEVSFKM 96

RESULT 6
Q9HPW4 PRELIMINARY; PRT; 77 AA.
ID Q9HPW4
AC Q9HPW4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
```

DT	01-MAR-2001	(TRENBLrel. 16, Last sequence update)
DT	01-JUN-2001	(TRENBLrel. 17, Last annotation update)
DE	Vng1446h.	
GN	VNG1446H.	
OS	Halobacterium sp. (strain NRC-1).	
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;	
OC	Halobacteriaceae; Halobacterium.	
OX	NCHI_TaxID=64091;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20504483; PubMed=11016950;	
RA	NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,	
RA	Sukla H.D., Lasky S.R., Ballig A.S., Thorsson V., Sorogna J.,	
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,	
RA	Leithausen B., Keller K., Cruz R., Banson M.J., Hough D.W.,	
RA	Madden D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,	
RA	Iisenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,	
RA	Ayam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,	
RA	Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;	
RT	"Genome sequence of Halobacterium species NRC-1";	
RL	Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).	
DR	EMBL; AE005061; AA019753.1; ..	
KW	Complete proteome.	
SQ	SEQUENCE 77 AA; 8378 MW; C3C59B65E8A8306 CRC64;	
Query Match 31.2%; Score 101; DB 17; Length 77;		
Best Local Similarity 31.1%; Pred.No. 0.00044;		
Matches 19; Conservative 22; Mismatches 20; Indels 0; Gaps 0;		
QY	4	IIGTSPTSQCQAABAVQRARDSVDDTRVARVIEQDMANVSAGKITVRKLEVSFWRPA 63
Db		: : : : : : : : : : : :
QY	64	Q 64
Db		77 Q 77
RESULT 7		
Q8RAAF	ID	Q8RAAF PRELIMINARY; PRG; 110 AA.
AC	Q8RAAF	AC
DT	01-JUN-2002	(TRENBLrel. 21, Created)
DT	01-JUN-2002	(TRENBLrel. 21, Last sequence update)
DT	01-OCT-2002	(TRENBLrel. 22, Last annotation update)
DE	Hypothetical protein TRE1318.	
GN	TRE1318.	
OS	Thermoanaerobacter tengcongensis.	
OC	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;	
OC	Thermoanaerobacteriaceae; Thermoanaerobacter.	
OX	NCHI_TaxID=119072;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	STRAIN=MB4 / JCM 11007;	
RX	MEDLINE=211992816; PubMed=11977336;	
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,	
RA	Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,	
RA	Tan H., Chen R., Wang J., Yu J., Yang H.;	
RT	"A complete sequence of T. tengcongensis genome.";	
RL	Genome Res. 12:689-700(2002).	
DR	EMBL; AE013092; AAM24542.1; ..	
KW	Hypothetical protein; Complete proteome.	
SQ	SEQUENCE 110 AA; 11862 MW; 43C0606A4B05F6DB CRC64;	
Query Match 28.4%; Score 92; DB 16; Length 110;		
Best Local Similarity 28.8%; Pred.No. 0.0071;		
Matches 17; Conservative 17; Mismatches 25; Indels 0; Gaps 0;		
QY	1	VIDIIGTSPSWCAAAEAQRARDSVDDRVARVIEQDMANVSAGKITVRKLEVSFK 59
Db		: : : : : : : : : : : :
Db	46	VLNVGDSTYSWDAIHKAVEAAKSPNISGLGVNQTANVGKIVEKANIQTAYR 104

```
Matches 15; Conservative 17; Mismatches 27; Indels 0; Gaps 0;
QY 2 IDIIGTSPTEQAAAEVQARDSDVDIRVARVIEQDMAVDSAGKIT-YRIKLEVSF 60
Db 10 IELGVSSKTSIEDANNALEAKSIQHLWFVVDTRGHIENGAVGHYQVTLKVGPR 68

RESULT 10
Q92X94 PRELIMINARY; PRT; 69 AA.
AC Q92X94;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein RB0068.
GN RB0068 OR SMB20068.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl Acad. Sci. U.S.A. 98:9889-9894 (2001).
DR EMBL; AL603642; CAC48468.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 69 AA; 7792 MW; F53BF80AE62C07D CRC64;

Query Match 20.4%; Score 66; DB 16; Length 69;
Best Local Similarity 27.9%; Pred. No. 3.8;
Matches 19; Conservative 17; Mismatches 14; Indels 18; Gaps 4;
QY 2 IDIIGTSPTEQAAAEVQARDSDVDIRVARVIEQDMAVDSAGKIT-YRI 52
Db 9 IELIGSPNSIDEAIEGASIRASKITRNLDFWFDQIR-GQIVN-----GKVARYQV 59

QY 53 KLEVSFPM 60
Db 60 VMKVGPR 67

RESULT 11
Q8XQP6 PRELIMINARY; PRT; 83 AA.
AC Q8XQP6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein Rsp1175.
GN RSP1175 OR RS05065.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salancoubat M., Genin S., Ariguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brotier P., Camus J.C., Cattolico L.,
RA Chandler M., Choigne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
```

```
RL Nature 415:497-502 (2002).
DR EMBL; AL646083; CAD18326.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 83 AA; 9140 MW; 1EAAD51D4FA4871A CRC64;

Query Match 20.1%; Score 65; DB 16; Length 83;
Best Local Similarity 28.8%; Pred. No. 6.1;
Matches 19; Conservative 20; Mismatches 23; Indels 4; Gaps 3;
QY 1 VIDIGTSPTEQAAAEVQARDSDVDIRVARVIEQDMAVDSAGKIT-YRIKLEVSF 58
Db 8 MIELVSSPESDAAIENAIARASEIHLVLDWFWEWETRGHLVD--GKIAHQVTLKVM 65

QY 59 KMRPAQ 64
Db 66 RLQEAQ 71

RESULT 12
Q8WVN1 PRELIMINARY; PRT; 501 AA.
AC Q8WVN1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glucose transporter 5.
GN SUC2A5.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSU=Small intestine;
RA Wood I.S., Castano-Merediz E.F., Dyer J., Shirazi-Beechey S.P.;
RT "Nutrient regulation of Glut5 in sheep intestine."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; AJ315928; CAC86964.1; -.
DR InterPro; IPR000005; HTHARAC.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub transporter.
DR InterPro; IPR003663; Sugar transp.
DR InterPro; IPR005829; Sug transporter.
DR Pfam; PF00083; sugar tr; 1.
DR PRINTS; PR00171; SUGTRNSPORT.
DR TIGRfams; TIGR00879; SP; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
DR PROSITE; PS00850; MFS; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Sugar transport; Transmembrane; Transport.
SQ SEQUENCE 501 AA; 55546 MW; 23F0E43E944C9AD5 CRC64;

Query Match 19.9%; Score 64.5; DB 6; Length 501;
Best Local Similarity 40.0%; Pred. No. 56;
Matches 20; Conservative 10; Mismatches 15; Indels 5; Gaps 3;
QY 13 EQAAAEVQARDSDVDIRVARVIEQDMAVDSAGKIT-YRIKLEVSF 61
Db 227 EEAARALRLRGWHDVDAIEIILBEDRAEKAAGFISV-LKL---FKMR 272

RESULT 13
Q962B3 PRELIMINARY; PRT; 1030 AA.
AC Q962B3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Aminopeptidase N (EC 3.4.11.2).
```

```
GN APN2.
OS Helicoverpa armigera (Cotton bollworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Heliothinae; Helicoverpa.
OX NCBI_TaxID=29058;
RN [1]
RP SEQUENCE FROM N.A.
RA Angelucci C., Akhurst R.J., East P.;
RT "Helicoverpa armigera aminopeptidase N gene APN2.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY038608; AAK85539.1; -.
DR MEROPS; W01.013; -.
DR InterPro; IPR001930; Ala peptase.
DR InterPro; IPR006025; Zn_MTPeptidse.
DR Pfam; PF01433; Peptidase M1; 1.
DR PRINTS; PR00756; ALADIPTASE.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Aminopeptidase; Hydrolase.
SQ SEQUENCE 1030 AA; 114505 MW; 258A43E74EFBFA7 CRC64;

Query Match 19.9%; Score 64.5; DB 5; Length 1030;
Best Local Similarity 30.4%; Pred. No. 1.3e+02;
Matches 14; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 11 SWEQAAAEAVQARDSDVDIRVARVIEQDVAVDSAGKITVRIKLEY 56
Db 610 NWEELAA-ALSRDHTAIHLNRAQIVDDVFALMRSGQITVRLGFKV 654

RESULT 14
Q8ZOH1 PRELIMINARY; PRT; 96 AA.
AC Q8ZOH1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional regulator.
GN ASL0125.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2159285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shampo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
PL DNA Res. 8:205-213(2001).
DR EMBL; AP003581; BAB77649.1; -.
DR InterPro; IPR001387; HTH_3.
DR Pfam; PF01381; HTH_3; 1.
DR SMART; SMO0530; HTH_XRE; 1.
KW Complete proteome.
SQ SEQUENCE 96 AA; 10897 MW; AC99D93E3B06C901 CRC64;

Query Match 19.8%; Score 64; DB 16; Length 96;
Best Local Similarity 26.7%; Pred. No. 9.4;
Matches 16; Conservative 10; Mismatches 28; Indels 6; Gaps 1;

QY 7 TSPTSWEQAAAEAVQARDSDVDIRVARVIEQDVAVDSAGKITVRIKLEVSFKMRPAQPR 66
Db 26 TSVRWAEGRTEPTWPKKYVDALRIFKCTPEELASASEKSINQR-----HKKPGRPR 79

RESULT 15
Q75741 PRELIMINARY; PRT; 206 AA.
ID Q75741
AC Q75741;
```

```
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nef (Negative factor) (P-protein) (27 kDa protein)
DE (Fragment).
GN NEF.
OS Human immunodeficiency virus 1.
OC Viruses; Retrod viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Mariani R., Kirchhoff F., Greenough T.C., Sullivan J.L.,
RA Desrochers R.C., Skowronski J.;
RT "High frequency of defective nef alleles in a long-term survivor with
nonprogressive HIV-1 infection.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
ACTIVITIES. IT SEEMS TO DOWN-REGULATE THE CD4(T4) ANTIGEN (BY
SIMILARITY).
CC EMBL; U61813; AAB04732.1; -.
DR HSSP; P03406; IEFN.
DR InterPro; IPR001558; HIV_Nef.
DR Pfam; PF00469; F-protein; 1.
DR ProDom; PD000031; HIV_Nef; 1.
KW AIDS; GTP-binding; Lipo-protein; Myristate.
FT NON_TER 1
SQ SEQUENCE 206 AA; 23974 MW; 12410737A86F8F25 CRC64;

Query Match 19.8%; Score 64; DB 15; Length 206;
Best Local Similarity 31.0%; Pred. No. 23;
Matches 18; Conservative 12; Mismatches 18; Indels 10; Gaps 2;

QY 18 EAVQARDSDVDIR-VASVIEQDVAVDSAGKITVRI-----KLEVSFKMRPAQP 65
Db 18 ERMRAEPAADKVRVSRVLEKHGAISSNTAANNVDCAWLEAQEEEVSPFRQVPP 75

Search completed: November 21, 2003, 16:08:10
Job time : 5.24762 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 5.7517 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 TDAATLAQAGNFERISGDL.....VOYSRADDEQQALSSQMGF 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	95	18 AAW32444	Mycobacterium tube
2	466	100.0	95	18 AAW32376	Mycobacterium tube
3	466	100.0	95	19 AAW81747	M. tuberculosis im
4	466	100.0	95	19 AAW64321	Mycobacterium tube
5	466	100.0	95	20 AAY32097	Mycobacterium tube
6	466	100.0	95	20 AAY39118	M. tuberculosis an
7	466	100.0	95	20 AAY38981	M. tuberculosis re
8	466	100.0	95	23 AAE29717	Mycobacterium tube
9	466	100.0	95	23 AAE17581	Mycobacterium spec

10	466	100.0	100	19 AAW81706	M. tuberculosis im
11	466	100.0	100	19 AAW64339	Mycobacterium tube
12	466	100.0	100	20 AAY39136	M. tuberculosis an
13	466	100.0	100	20 AAY38993	M. tuberculosis re
14	466	100.0	100	22 AAB35218	M. tuberculosis RV3
15	466	100.0	100	22 AAB19845	Mycobacterium tube
16	466	100.0	100	23 ABU05988	M. tuberculosis an
17	466	100.0	100	23 ARG30969	Mycobacterium tube
18	466	100.0	358	23 AAU74591	Antigenic fusion p
19	466	100.0	802	19 AAW81746	M. tuberculosis fu
20	466	100.0	802	19 AAW64379	Mycobacterium anti
21	466	100.0	802	20 AAY32063	Mycobacterium tube
22	466	100.0	802	20 AAY39224	M. tuberculosis fu
23	466	100.0	802	20 AAY39176	M. tuberculosis fu
24	466	100.0	802	20 AAY39081	M. tuberculosis fus
25	466	100.0	802	20 AAY39033	M. tuberculosis fu
26	466	100.0	802	23 AAU74592	Antigenic fusion p
27	466	100.0	983	22 AAW01901	M. tuberculosis TB
28	462	99.1	100	20 AAY03705	M. tuberculosis LH
29	392	84.1	80	18 AAW32454	Mycobacterium tube
30	392	84.1	80	18 AAW32386	Mycobacterium tube
31	392	84.1	80	19 AAW81707	M. tuberculosis im
32	392	84.1	80	19 AAW64340	Mycobacterium tube
33	392	84.1	80	20 AAY39137	M. tuberculosis an
34	392	84.1	80	20 AAY38994	M. tuberculosis re
35	217	46.6	49	20 AAY03706	M. tuberculosis LH
36	205	44.0	42	20 AAY03707	M. tuberculosis LH
37	173	37.1	100	23 ABU05361	M. tuberculosis an
38	137	29.4	28	20 AAY03712	M. tuberculosis LH
39	122	26.2	28	18 AAW32460	Mycobacterium tube
40	122	26.2	28	19 AAW81698	M. tuberculosis im
41	122	26.2	28	20 AAY39128	M. tuberculosis an
42	121	26.0	27	18 AAW32458	Mycobacterium tube
43	121	26.0	27	19 AAW81696	M. tuberculosis im
44	121	26.0	27	20 AAY39126	M. tuberculosis an
45	118	25.3	27	18 AAW32457	Mycobacterium tube

ALIGNMENTS

```
RESULT 1
AAW32444
ID AAW32444 standard; Protein; 95 AA.
XX
AC AAW32444;
DT 09-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen Tb38-1.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
XX
XX WO9709428-A2.
XX
PD 13-MAR-1997.
XX
XX
XX 30-AUG-1996; 96WO-US14674.
XX
PR 12-JUL-1996; 96US-0680574.
PR 01-SEP-1995; 95US-0523436.
PR 22-SEP-1995; 95US-0533634.
PR 22-MAR-1996; 96US-0620874.
PR 05-JUN-1996; 96US-0659683.
XX
(CORI-) CORIXA CORP.
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
PI Twardzik DR, Vedvick TH;
XX
```

DR WPI; 1997-192903/17.
 XX N-PSDB; AAT91509.
 PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also
 PT for diagnosis
 XX Example 3; Page 124; 168pp; English.
 XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX Sequence 95 AA;
 SQ Query Match 100.0%; Score 466; DB 18; Length 95;
 Best Local Similarity 100.0%; Pred. No. 2.1e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFEAAANKQ 60
 DB 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFEAAANKQ 60
 QY 61 KQELDEISTNIRAGVQYGRADDEEQQALSSQMGF 95
 DB 61 KQELDEISTNIRAGVQYGRADDEEQQALSSQMGF 95
 RESULT 2
 AAW32376
 ID AAW32376 standard; Protein; 95 AA.
 XX AAW32376;
 AC AAW32376;
 DT 13-JAN-1998 (first entry)
 XX Mycobacterium tuberculosis antigen Tb38-1.
 DE Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 XX skin testing; M.tuberculosis.
 XX Mycobacterium tuberculosis.
 OS WO9709429-A2.
 XX 13-MAR-1997.
 PD 13-MAR-1997.
 XX 30-AUG-1996; 96WO-US14675.
 XX 12-JUL-1996; 96US-0680573.
 PR 01-SEP-1995; 95US-0523435.
 PR 22-SEP-1995; 95US-0532136.
 PR 22-MAR-1996; 96US-0620280.
 PR 05-JUN-1996; 96US-0658800.
 XX (CORI-) CORIXA CORP.
 PA Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
 PI Twardzik DR, Vedvick TH;
 XX WPI; 1997-192904/17.
 DR N-PSDB; AAT91445.
 XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
 PT - useful for diagnosis of M. tuberculosis infection
 XX Example 3; Page 136; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX Sequence 95 AA;
 SQ Query Match 100.0%; Score 466; DB 18; Length 95;
 Best Local Similarity 100.0%; Pred. No. 2.1e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFEAAANKQ 60
 DB 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFEAAANKQ 60
 QY 61 KQELDEISTNIRAGVQYGRADDEEQQALSSQMGF 95
 DB 61 KQELDEISTNIRAGVQYGRADDEEQQALSSQMGF 95
 RESULT 3
 AAW81747
 ID AAW81747 standard; Protein; 95 AA.
 XX AAW81747;
 AC AAW81747;
 DT 27-JAN-1999 (first entry)
 XX M. tuberculosis immunogenic polypeptide Tb38-1.
 DE Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 XX vaccine; pharmaceutical; infection; diagnosis.
 XX Mycobacterium tuberculosis.
 OS WO9816646-A2.
 XX 23-APR-1998.
 PD 07-OCT-1997; 97WO-US18293.
 XX 13-MAR-1997; 97US-0818112.
 PR 11-OCT-1996; 96US-0730510.
 XX (CORI-) CORIXA CORP.
 PA Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64491.
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX Example 3b; Page 117; 230pp; English.
 PS This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
 CC for inducing protective immunity against tuberculosis (TB). This sequence
 CC can be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis.
 XX Sequence 95 AA;


```

Query Match      100.0%; Score 466; DB 19; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60

QY 61 KOELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95

RESULT 4
AAW64321
ID AAW64321 standard; Peptide; 95 AA.
AC AAW64321;
DT 09-NOV-1998 (first entry)
DE Mycobacterium tuberculosis antigen Tb38-1 peptide.
XX Tuberculosis; infection; diagnosis; antigen; Tb38-1.
KW Mycobacterium tuberculosis strain H37Rv.
OS WO9816645-A2.
FN 23-APR-1998.
PD 07-OCT-1997; 97WO-US18214.
PF 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR WPI; 1998-251292/22.
DR N-PSDB; AAW44384.
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX Example 3; Page 123; 250pp; English.
XX This is an antigenic portion of Mycobacterium tuberculosis antigen
CC Tb38-1. A DNA sequence (see AAW44384) coding for antigen Tb38-1 was
CC isolated from a M. tuberculosis strain H37Rv expression library
CC using sera from patients having pulmonary or pleural tuberculosis.
CC The invention relates to compositions and methods for diagnosing
CC tuberculosis. It provides polypeptides (see AAW64291-W64379)
CC comprising an antigenic portion of a soluble M. tuberculosis
CC antigen, or an immunogenic portion of an M. tuberculosis antigen,
CC as well as DNA sequences encoding such polypeptides, recombinant
CC expression vectors and transformed or transfected host cells. Also
CC claimed are methods and diagnostic kits for detecting M.
CC tuberculosis infection in a patient using these polypeptides,
CC antibodies or oligonucleotide probes and primers, for the diagnosis
CC of tuberculosis.
XX
SQ Sequence 95 AA;
Query Match      100.0%; Score 466; DB 19; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60

QY 61 KOELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95

RESULT 6
AAW39118
ID AAW39118 standard; Peptide; 95 AA.
XX

```

```

Db 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
QY 61 KOELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95

RESULT 5
AAW32097
ID AAW32097 standard; Protein; 95 AA.
XX AAW32097;
XX 17-JAN-2000 (first entry)
DT Mycobacterium tuberculosis antigen Tb38-1.
DE Tuberculosis; antigen; fusion protein; Tb38-1; diagnosis; therapy;
XX vaccine; immunogen.
XX Mycobacterium tuberculosis.
XX WO9951748-A2.
FN 14-OCT-1999.
PD 07-APR-1999; 99WO-US07717.
PF 07-APR-1998; 98US-0056556.
PR 30-DEC-1998; 98US-0223040.
XX (CORI-) CORIXA CORP.
XX Skeiky YAW, Alderson M, Campos-Neto A;
XX WPI; 1999-601610/51.
XX New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -
XX Claim 1; Fig 4D; 83pp; English.
XX This sequence represents the Mycobacterium tuberculosis antigen
CC Tb38-1. The invention provides fusion proteins (see AAY32059-71)
CC containing at least 2 M. tuberculosis antigens such as Tb38-1, e.g.
CC TBP-2 (see AAY32063) and a TBP9-Tb38-1 fusion. The new fusion
CC proteins are useful as vaccines for preventing tuberculosis
CC (claimed), for diagnosis (via in vitro assays or intradermal skin
CC tests for detection of anti-M. tuberculosis antibodies), monitoring
CC of disease progression, and treatment of tuberculosis. They are
CC more effective immunogens than mixtures of the individual protein
CC components.
XX
SQ Sequence 95 AA;
Query Match      100.0%; Score 466; DB 20; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60

QY 61 KOELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95

RESULT 6
AAW39118
ID AAW39118 standard; Protein; 95 AA.
XX

```

AC AAY39118;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis antigen Tb38-1 amino acid sequence.
 XX
 DE Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03268.
 XX
 PR 05-MAY-1998; 98US-0072967.
 XX
 PR 18-FEB-1998; 98US-0025197.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX WPI; 1999-527409/44.
 DR
 DR New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Example 3; Page 113; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T. B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AAY39249 to AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.
 XX
 SQ Sequence 95 AA;

Query Match 100.0%; Score 466; DB 20; Length 95;
 Best Local Similarity 100.0%; Pred. No. 2.1e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEANKQ 60
 DB 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEANKQ 60
 QY 61 KQELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95
 DB 61 KQELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95

RESULT 7
 AAY38981
 ID AAY38981 standard; Protein; 95 AA.
 XX
 AC AAY38981;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis recombinant antigen protein Tb38-1.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX

OS Mycobacterium tuberculosis.
 XX
 PN WO9942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03265.
 XX
 PR 05-MAY-1998; 98US-0072596.
 XX
 PR 18-FEB-1998; 98US-0024753.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX WPI; 1999-527416/44.
 DR
 DR N-FSDB; AAZ19082.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis
 XX
 PS Example 3; Page 159; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX
 SQ Sequence 95 AA;

Query Match 100.0%; Score 466; DB 20; Length 95;
 Best Local Similarity 100.0%; Pred. No. 2.1e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEANKQ 60
 DB 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEANKQ 60
 QY 61 KQELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95
 DB 61 KQELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95

RESULT 8
 AAE29717
 ID AAE29717 standard; Protein; 95 AA.
 XX
 AC AAE29717;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium tuberculosis Tb38-1 antigenic protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; Tb38-1 antigen;
 KW MTb11.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200272792-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 13-MAR-2002; 2002WO-US08223.
 XX
 PR 13-MAR-2001; 2001US-275837P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Brannon M, Guderian J;
 XX

```

DR WPI; 2002-759844/82.
DR N-PSDB; AAD47094.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA,
XX LeIF, M15 or 5H polynucleotide, useful as vaccine to elicit protective
PT immunity against pathogenic microorganisms e.g. Leishmania and
PT Mycobacterium tuberculosis
XX
XX
PS Disclosure; Page 107; 155pp; English.
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from T9A, LeIF, M15, and 5H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides,
CC as in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC M. tuberculosis Tb38-1 antigenic protein. Tb38-1 is also referred to
CC as MTb11 or 38-1.
XX
SQ Sequence 95 AA;
Query Match 100.0%; Score 466; DB 23; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
DB 1 TDAATLAQAGNFERISGDLKTIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
QY 61 KQELDEISTNIRQAGVQVSRADDEEQQALSSQMGF 95
DB 61 KQELDEISTNIRQAGVQVSRADDEEQQALSSQMGF 95
RESULT 9
AAE17581
ID AAE17581 standard; Protein; 95 AA.
XX
AC AAE17581;
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species Tb38-1 (Mtbi1; 38-1) protein.
XX
KW Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; Tb38-1; Mtbi1; 38-1 protein.
XX
OS Mycobacterium sp.
XX
FN WO2001198460-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US19959.
XX
PR 20-JUN-2000; 2000US-0597796.
PR 01-FEB-2001; 2001US-265737P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Alderson M;
XX
XX WPI; 2002-147798/19.
DR N-PSDB; AAD28352.
XX
PT Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a

```

```

PT subject -
XX
PS Claim 9; Page 123; 136pp; English.
XX
CC The present invention relates to fusion proteins containing at least
CC two Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected
CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC Tb38-1 (Mtbi1; 38-1) protein.
XX
SQ Sequence 95 AA;
Query Match 100.0%; Score 466; DB 23; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
DB 1 TDAATLAQAGNFERISGDLKTIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
QY 61 KQELDEISTNIRQAGVQVSRADDEEQQALSSQMGF 95
DB 61 KQELDEISTNIRQAGVQVSRADDEEQQALSSQMGF 95
RESULT 10
AAW81706
ID AAW81706 standard; Protein; 100 AA.
XX
AC AAW81706;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide Tb38-IN.
XX
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
FN WO9816646-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; 97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-261042/23.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used

```

PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX
 PS Example 3B; Page 138-139; 230pp; English.
 XX
 CC This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
 CC for inducing protective immunity against tuberculosis (TB). This
 CC sequence can be formulated into vaccines and/or pharmaceutical
 CC compositions for immunising against M. tuberculosis infection or may
 CC be used for the diagnosis of tuberculosis.
 XX
 XX Sequence 100 AA;
 SQ
 Query Match 100.0%; Score 466; DB 19; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.2e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 TDAATLAQEAAGNERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
 Db 6 TDAATLAQEAAGNERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 65
 QY 61 KQELDEISTNIRQAGVQYSPADEEQQALSSQMGF 95
 Db 66 KQELDEISTNIRQAGVQYSPADEEQQALSSQMGF 100
 RESULT 11
 AAW64339
 ID AAW64339 standard; Protein; 100 AA.
 XX
 AC AAW64339;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Tb38-IN.
 XX
 KW Tuberculosis; infection; diagnosis; antigen; Tb37-FL.
 XX
 OS Mycobacterium tuberculosis strain H37RV.
 XX
 FN WO9816645-A2.
 XX
 XX 23-APR-1998.
 XX
 XX 07-OCT-1997; 97WO-US18214.
 XX
 PR 13-MAR-1997; 97US-0818111.
 PR 11-OCT-1996; 96US-0729622.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1998-251292/22.
 XX
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and diagnosis of tuberculosis
 XX
 PS Example 3; Page 145; 250pp; English.
 XX
 CC This polypeptide comprises a partial sequence of Mycobacterium
 CC tuberculosis antigen Tb38-IN. It is encoded by genomic DNA isolated
 CC from a M. tuberculosis strain H37RV genomic library using a probe
 CC derived from clone Tb38-1 (see AAW44384). The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of
 CC a soluble M. tuberculosis antigen, or an immunogenic portion of an
 CC M. tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic

CC kits for detecting M. tuberculosis infection in a patient using
 CC these polypeptides, antibodies or oligonucleotide probes and
 CC primers, for the diagnosis of tuberculosis.
 XX
 PS Sequence 100 AA;
 SQ
 Query Match 100.0%; Score 466; DB 19; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.2e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 TDAATLAQEAAGNERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
 Db 6 TDAATLAQEAAGNERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 65
 QY 61 KQELDEISTNIRQAGVQYSPADEEQQALSSQMGF 95
 Db 66 KQELDEISTNIRQAGVQYSPADEEQQALSSQMGF 100
 RESULT 12
 AAY39136
 ID AAY39136 standard; Protein; 100 AA.
 XX
 AC AAY39136;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis antigen Tb38-IN amino acid sequence.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03268.
 XX
 PR 05-MAY-1998; 98US-0072967.
 PR 18-FEB-1998; 98US-0025137.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R,
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1999-527409/44.
 XX
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Example 3; Page 133-134; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.
 XX
 XX Sequence 100 AA;
 SQ
 Query Match 100.0%; Score 466; DB 20; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.2e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 60
 |||||
 Db 6 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 65
 |||||
 QY 61 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
 |||||
 Db 66 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
 |||||

RESULT 13

AAV38993
 ID AAV38993 standard; Protein; 100 AA.

XX AC AAV38993;
 XX DT 05-NOV-1999 (first entry)
 XX DE M. tuberculosis recombinant antigen protein Tb38-IN.
 XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 XX KW vaccine; immunity.

XX OS Mycobacterium tuberculosis.
 XX PN WO9942118-A2.
 XX PD 26-AUG-1999.
 XX PF 17-FEB-1999; 98WO-US03265.
 XX PR 05-MAY-1998; 98US-0072596.
 XX PR 18-FEB-1998; 98US-0024753.
 XX PA (CORI-) CORIXA CORP.
 XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX DR WPI; 1999-527416/44.
 XX PT New polypeptide comprising antigenic portions of M. tuberculosis
 XX PS Example 3; Page 179; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX SQ Sequence 100 AA;

Query Match 100.0%; Score 466; DB 20; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.2e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 60
 |||||
 Db 6 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 65
 |||||
 QY 61 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
 |||||
 Db 66 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
 |||||

RESULT 14

AAV35218
 ID AAV35218 standard; Protein; 100 AA.

XX AC AAV35218;

XX

DT 24-APR-2001 (first entry)

DE M tuberculosis Rv3874 protein.

XX KW Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c;
 KW Rv1037c; Rv2348c; Rv2653c; Rv2654c; Rv3020c; Rv3444c;
 KW Rv445c; Rv3890c; Rv3891c; Rv3904c; Rv3905c.

XX OS Mycobacterium tuberculosis.
 XX PN WO200104151-A2.
 XX PD 18-JAN-2001.
 XX PF 13-JUL-2000; 2000WO-DK00398.
 XX PR 13-JUL-1999; 99DK-0001020.
 XX PR 15-JUL-1999; 99US-0144011.
 XX PA (STAT-) STATENS SERUM INST.
 XX PI Andersen P, Skjot R;
 XX DR WPI; 2001-091923/10.
 XX PT New polypeptide encoded by a member of the esat-6-gene family for
 PT immunizing against and diagnosis of tuberculosis .
 XX PS Example 2; Page 65; 80pp; English.

XX CC The present invention provides the protein and coding sequences for
 CC members of the esat-6 gene family from Mycobacterium tuberculosis. These
 CC proteins include Rv0287, Rv1036c, Rv1037c, Rv2348c, Rv2653c,
 CC Rv2654c, Rv3020c, Rv3444c, Rv3890c, Rv3891c, Rv3904c and
 CC Rv3905c. These can be used to produce vaccines against, and in the
 CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of
 CC the proteins of the invention.
 XX SQ Sequence 100 AA;

Query Match 100.0%; Score 466; DB 22; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.2e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 60
 |||||
 Db 6 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 65
 |||||
 QY 61 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
 |||||
 Db 66 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
 |||||

RESULT 15

AAV19845
 ID AAV19845 standard; Protein; 100 AA.

XX AC AAV19845;
 XX DT 05-MAR-2001 (first entry)
 XX DE Mycobacterium tuberculosis protein MTEN4.
 XX KW MTEN4; tuberculosis; BCG; vaccine; infection; diagnosis.
 XX OS Mycobacterium tuberculosis.
 XX PN WO200006157-A1.
 XX PD 09-NOV-2000.
 XX PF 04-MAY-2000; 2000WO-US12257.

```

XX 04-MAY-1999; 99US-0132505.
XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.
XX Gennaro ML;
XX WPI: 2001-007153/01.
XX N-PSDB; AAA89038.
XX Novel polypeptide encoded by open reading frames present in
PT Mycobacterium tuberculosis genome and not by the BCG strain of M.
PT bovis, useful as vaccine and for diagnosing tuberculosis infection
XX Claim 11; Fig 1; 35pp; English.
XX The present sequence is that of the Mycobacterium tuberculosis
CC MTEN4 protein. This is 1 of 8 proteins, i.e. MTEN1-8 (see
CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)
CC identified as being present in the genome of M. tuberculosis but
CC absent from the genome of the BCG strain of Mycobacterium bovis.
CC MTEN1-8 represent reagents that are useful in discriminating between
CC M. tuberculosis and BCG and, in particular, for diagnostic methods
CC which discriminate between exposure of a subject to M. tuberculosis
CC and vaccination with BCG. The invention features these MTEN
CC polypeptides, functional fragments of them, DNA encoding them,
CC vectors, transformed cells, and diagnostic, therapeutic, and
CC prophylactic (vaccine) methods, including genetic vaccination
XX methods.
XX SQ Sequence 100 AA;
Query Match 100.0%; Score 466; DB 22; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDAATLAQEAAGNFERISGDKTQIDOVETAGSLQGWREGAGTAAQAQAVVRQEAANKQ 60
Db 6 TDAATLAQEAAGNFERISGDKTQIDOVETAGSLQGWREGAGTAAQAQAVVRQEAANKQ 65
Qy 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
Db 66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

```

Search completed: November 21, 2003, 16:03:12
Job time : 5.7517 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 Search time 2.0034 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 TDAATLAQAGNFRISGDL.....VOYSRADDEQQALSSQMGF 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/6C.COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/6D.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	95	3	US-08-818-112-88
2	466	100.0	95	4	US-08-818-111-89
3	466	100.0	95	4	US-09-056-556-88
4	466	100.0	95	4	US-09-072-536-89
5	466	100.0	100	3	US-08-818-112-115
6	466	100.0	100	4	US-08-818-111-110
7	466	100.0	100	4	US-09-056-556-115
8	466	100.0	100	4	US-09-072-536-110
9	466	100.0	802	4	US-09-056-556-214
10	466	100.0	802	4	US-09-072-536-209
11	466	100.0	802	4	US-09-072-536-346
12	462	99.1	100	4	US-09-116-492A-5
13	392	84.1	180	3	US-08-818-112-117
14	392	84.1	80	4	US-08-818-111-112
15	392	84.1	80	4	US-09-056-556-117
16	392	84.1	80	4	US-09-072-536-112
17	217	46.6	49	4	US-08-116-492A-6
18	205	44.0	42	4	US-09-116-492A-7
19	173	37.1	100	4	US-09-116-492A-28
20	137	29.4	28	4	US-09-116-492A-12
21	122	26.2	28	3	US-08-818-112-98
22	122	26.2	28	4	US-09-056-556-98
23	121	26.0	27	3	US-08-818-112-96
24	121	26.0	27	4	US-08-056-556-96
25	118	25.3	27	3	US-08-818-112-95
26	118	25.3	27	3	US-08-818-112-97
27	118	25.3	27	4	US-09-056-556-95

28	118	25.3	27	4	US-09-056-556-97	Sequence 97, Appl
29	117	25.1	28	3	US-08-818-112-93	Sequence 93, Appl
30	117	25.1	28	4	US-09-056-556-93	Sequence 93, Appl
31	100	21.5	20	4	US-09-116-492A-8	Sequence 8, Appl
32	87.5	18.8	97	4	US-09-073-009-109	Sequence 109, App
33	87.5	18.8	97	4	US-09-073-009-143	Sequence 143, App
34	85	18.2	16	3	US-08-818-112-94	Sequence 94, Appl
35	85	18.2	16	4	US-09-056-556-94	Sequence 94, Appl
36	81.5	17.5	955	1	US-08-006-676B-1	Sequence 1, Appl
37	81.5	17.5	955	1	US-08-282-845-2	Sequence 1, Appl
38	81.5	17.5	955	2	US-08-428-414A-3	Sequence 3, Appl
39	81.5	17.5	955	5	PC-US94-00324-1	Sequence 3, Appl
40	79	17.0	16	4	US-09-116-492A-10	Sequence 10, Appl
41	78	16.7	16	4	US-09-116-492A-9	Sequence 9, Appl
42	78	16.7	16	4	US-09-116-492A-13	Sequence 13, Appl
43	77	16.5	16	4	US-09-116-492A-11	Sequence 11, Appl
44	77	16.5	336	4	US-09-252-991A-24656	Sequence 24656, A
45	75	16.1	1147	1	US-08-144-121-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-08-818-112-88
; Sequence 88, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-88

Query Match 100.0%; Score 466; DB 3; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60

QY 61 KQELDEISTNIRQAGVQYSRADEEQQAALSSQMGF 95
Db 61 KQELDEISTNIRQAGVQYSRADEEQQAALSSQMGF 95

RESULT 2

US-08-818-111-89
; Sequence 89, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/818,111
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-818-111-89
Query Match 100.0%; Score 466; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 2,1e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
QY 61 KQELDEISTNIRQAGVQYSRADEEQQAALSSQMGF 95
Db 61 KQELDEISTNIRQAGVQYSRADEEQQAALSSQMGF 95

US-08-818-111-89
Query Match 100.0%; Score 466; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 2,1e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
QY 61 KQELDEISTNIRQAGVQYSRADEEQQAALSSQMGF 95
Db 61 KQELDEISTNIRQAGVQYSRADEEQQAALSSQMGF 95

RESULT 3

US-09-056-556-88
; Sequence 88, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/056,556
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-056-556-88

Query Match 100.0%; Score 466; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 2,1e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
QY 61 KQELDEISTNIRQAGVQYSRADEEQQAALSSQMGF 95
Db 61 KQELDEISTNIRQAGVQYSRADEEQQAALSSQMGF 95

RESULT 4

US-09-072-596-89
; Sequence 89, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk


```

; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-112-115

Query Match
Best Local Similarity 100.0%; Score 466; DB 3; Length 100;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
DB 6 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 65

QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
DB 66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 6
US-08-818-111-110
; Sequence 110, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Read, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-111-110

Query Match
Best Local Similarity 100.0%; Score 456; DB 4; Length 100;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
DB 6 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 65

QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
DB 66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 5
US-08-818-112-115
; Sequence 115, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Read, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031

```


Query Match	100.0%	Score 466;	DB 4;	Length 802;
Best Local Similarity	100.0%	Pred. No. 3.5e-45;		
Matches	95;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps	0;			
QY	1	TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ	60	
Db	429	TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ	488	
QY	61	KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF	95	
Db	489	KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF	523	
RESULT 11				
US-09-072-596-346				
Sequence 346,	Application US/09072596			
Patent No. 6458366				
GENERAL INFORMATION:				
APPLICANT:	Reed, Steven G.			
APPLICANT:	Steiky, Yasir A.W.			
APPLICANT:	Dillon, Davin C.			
APPLICANT:	Campos-Neto, Antonia			
APPLICANT:	Houghton, Raymond			
APPLICANT:	Vedvick, Thomas S.			
APPLICANT:	Twardzik, Daniel R.			
APPLICANT:	Lodes, Michael J.			
APPLICANT:	Hendrickson, Ronald C.			
TITLE OF INVENTION:	COMPOUNDS AND METHODS FOR DIAGNOSIS OF			
NUMBER OF SEQUENCES:	350			
CORRESPONDENCE ADDRESS:				
ADDRESSEE:	SEED AND BERRY LLP			
STREET:	6300 Columbia Center, 701 Fifth Avenue			
CITY:	Seattle			
STATE:	Washington			
COUNTRY:	USA			
ZIP:	98104-7092			
COMPUTER READABLE FORM:				
MEDIUM TYPE:	Floppy disk			
COMPUTER:	IBM PC compatible			
OPERATING SYSTEM:	PC-DOS/MS-DOS			
SOFTWARE:	Patent In Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:				
APPLICATION NUMBER:	US/09/072,596			
FILING DATE:	05-MAY-1998			
CLASSIFICATION:				
ATTORNEY/AGENT INFORMATION:				
NAME:	Maki, David J.			
REGISTRATION NUMBER:	31,392			
REFERENCE/DOCKET NUMBER:	210121.417C9			
TELECOMMUNICATION INFORMATION:				
TELEPHONE:	(206) 622-4900			
TELEFAX:	(206) 682-6031			
INFORMATION FOR SEQ ID NO:	346:			
SEQUENCE CHARACTERISTICS:				
LENGTH:	802 amino acids			
TYPE:	amino acid			
STRANDEDNESS:	single			
TOPOLOGY:	linear			
MOLECULE TYPE:	protein			
US-09-072-596-346				
Query Match	100.0%	Score 466;	DB 4;	Length 802;
Best Local Similarity	100.0%	Pred. No. 3.5e-45;		
Matches	95;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps	0;			
QY	1	TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ	60	
Db	429	TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ	488	
QY	61	KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF	95	
Db	489	KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF	523	
RESULT 10				
US-09-072-596-209				
Sequence 209,	Application US/09072596			
Patent No. 6458366				
GENERAL INFORMATION:				
APPLICANT:	Reed, Steven G.			
APPLICANT:	Steiky, Yasir A.W.			
APPLICANT:	Dillon, Davin C.			
APPLICANT:	Campos-Neto, Antonia			
APPLICANT:	Houghton, Raymond			
APPLICANT:	Vedvick, Thomas S.			
APPLICANT:	Twardzik, Daniel R.			
APPLICANT:	Lodes, Michael J.			
APPLICANT:	Hendrickson, Ronald C.			
TITLE OF INVENTION:	COMPOUNDS AND METHODS FOR DIAGNOSIS OF			
NUMBER OF SEQUENCES:	350			
CORRESPONDENCE ADDRESS:				
ADDRESSEE:	SEED AND BERRY LLP			
STREET:	6300 Columbia Center, 701 Fifth Avenue			
CITY:	Seattle			
STATE:	Washington			
COUNTRY:	USA			
ZIP:	98104-7092			
COMPUTER READABLE FORM:				
MEDIUM TYPE:	Floppy disk			
COMPUTER:	IBM PC compatible			
OPERATING SYSTEM:	PC-DOS/MS-DOS			
SOFTWARE				

Query Match	100.0%	Score 466;	DB 4;	Length 802;
Best Local Similarity	100.0%	Pred. No. 3.5e-45;		
Matches	95;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps	0;			
QY	1	TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGQWRGAAGTAAQAQAAVVRFOEAANKQ	60	
Db	429	TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGQWRGAAGTAAQAQAAVVRFOEAANKQ	488	
QY	61	KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF	95	
Db	489	KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF	523	
RESULT 11				
US-09-072-596-346				
Sequence 346,	Application US/09072596			
Patent No. 6458366				
GENERAL INFORMATION:				
APPLICANT:	Reed, Steven G.			
APPLICANT:	Steiky, Yasir A.W.			
APPLICANT:	Dillon, Davin C.			
APPLICANT:	Campos-Neto, Antonia			
APPLICANT:	Houghton, Raymond			
APPLICANT:	Vedvick, Thomas S.			
APPLICANT:	Twardzik, Daniel R.			
APPLICANT:	Lodes, Michael J.			
APPLICANT:	Hendrickson, Ronald C.			
TITLE OF INVENTION:	COMPOUNDS AND METHODS FOR DIAGNOSIS OF			
NUMBER OF SEQUENCES:	350			
CORRESPONDENCE ADDRESS:				
ADDRESSEE:	SEED AND BERRY LLP			
STREET:	6300 Columbia Center, 701 Fifth Avenue			
CITY:	Seattle			
STATE:	Washington			
COUNTRY:	USA			
ZIP:	98104-7092			
COMPUTER READABLE FORM:				
COMPUTER:	IBM PC compatible			
OPERATING SYSTEM:	PC-DOS/MS-DOS			
SOFTWARE:	Patent In Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:				
APPLICATION NUMBER:	US/09/072,596			
FILING DATE:	05-MAY-1998			
CLASSIFICATION:				
ATTORNEY/AGENT INFORMATION:				
NAME:	Maki, David J.			
REGISTRATION NUMBER:	31,392			
REFERENCE/DOCKET NUMBER:	210121.417C9			
TELECOMMUNICATION INFORMATION:				
TELEPHONE:	(206) 622-4900			
TELEFAX:	(206) 682-6031			
INFORMATION FOR SEQ ID NO:	346:			
SEQUENCE CHARACTERISTICS:				
LENGTH:	802 amino acids			
TYPE:	amino acid			
STRANDEDNESS:	single			
TOPOLOGY:	linear			
MOLECULE TYPE:	protein			
US-09-072-596-346				
Query Match	100.0%	Score 466;	DB 4;	Length 802;
Best Local Similarity	100.0%	Pred. No. 3.5e-45;		
Matches	95;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps	0;			
QY	1	TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGQWRGAAGTAAQAQAAVVRFOEAANKQ	60	
Db	429	TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGQWRGAAGTAAQAQAAVVRFOEAANKQ	488	
QY	61	KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF	95	
Db	489	KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF	523	
RESULT 10				
US-09-072-596-209				
Sequence 209,	Application US/09072596			
Patent No. 6458366				
GENERAL INFORMATION:				
APPLICANT:	Reed, Steven G.			
APPLICANT:	Steiky, Yasir A.W.			
APPLICANT:	Dillon, Davin C.			
APPLICANT:	Campos-Neto, Antonia			
APPLICANT:	Houghton, Raymond			
APPLICANT:	Vedvick, Thomas S.			
APPLICANT:	Twardzik, Daniel R.			
APPLICANT:	Lodes, Michael J.			
APPLICANT:	Hendrickson, Ronald C.			
TITLE OF INVENTION:	COMPOUNDS AND METHODS FOR DIAGNOSIS OF			
NUMBER OF SEQUENCES:	350			
CORRESPONDENCE ADDRESS:				
ADDRESSEE:	SEED AND BERRY LLP			
STREET:	6300 Columbia Center, 701 Fifth Avenue			
CITY:	Seattle			
STATE:	Washington			
COUNTRY:	USA			
ZIP:	98104-7092			
COMPUTER READABLE FORM:				
COMPUTER:	IBM PC compatible			
OPERATING SYSTEM:	PC-DOS/MS-DOS			
SOFTWARE:	Patent In Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:				

Db 489 KQELDEISTNIRAGVQYSRADDEQQQALSSQMGF 523

RESULT 12

US-09-116-492A-5

Sequence 5, Application US/09116492A

Patent No. 6436409

GENERAL INFORMATION:

APPLICANT: GICQUEL, BRIGITTE

APPLICANT: BERTHET, FRANCOIS-XAVIER

APPLICANT: ANDERSEN, PETER

APPLICANT: RASMUSSEN, PETER B

TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS W

TITLE OF INVENTION: USING THE SAME

FILE REFERENCE: 0660-0137-27X

CURRENT APPLICATION NUMBER: US/09/116,492A

CURRENT FILING DATE: 1998-07-16

PRIOR APPLICATION NUMBER: 60/252,631

PRIOR FILING DATE: 1997-07-16

NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn version 3.1

SEQ ID NO 5

LENGTH: 100

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

US-09-116-492A-5

Query Match 99.1%; Score 462; DB 4; Length 100;

Best Local Similarity 98.9%; Pred. No. 6.5e-46;

Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TDAATLAQEAAGNERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 60

Db 6 TDAATLQEAAGNERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 65

QY 61 KQELDEISTNIRAGVQYSRADDEQQQALSSQMGF 95

Db 66 KQELDEISTNIRAGVQYSRADDEQQQALSSQMGF 100

RESULT 13

US-08-818-112-117

Sequence 117, Application US/08818112

Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Vedvick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY OF TUBERCULOSIS

TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,112

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C6

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 622-4900

INFORMATION FOR SEQ ID NO: 112:

SEQUENCE CHARACTERISTICS:

LENGTH: 80 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-818-111-112

Query Match 84.1%; Score 392; DB 4; Length 80;

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C6

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 117:

SEQUENCE CHARACTERISTICS:

LENGTH: 80 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-818-112-117

Query Match 84.1%; Score 392; DB 3; Length 80;

Best Local Similarity 100.0%; Pred. No. 5.3e-38;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 ISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQKQELDEISTNIRQAG 75

Db 1 ISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQKQELDEISTNIRQAG 60

QY 76 VOYSRADEEQQALSSQMGF 95

Db 61 VOYSRADEEQQALSSQMGF 80

RESULT 14

US-08-818-111-112

Sequence 112, Application US/08818111

Patent No. 6338852

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Vedvick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 148

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,111

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C6

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 112:

SEQUENCE CHARACTERISTICS:

LENGTH: 80 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-818-111-112

Query Match 84.1%; Score 392; DB 4; Length 80;

Search completed: November 21, 2003, 16:11:38
Job time : 2.0034 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 3.70522 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 PDATLAQENGRISGL.....VOYSADEEQQALSSQWGF 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US50_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US50_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	95	12	US-10-084-843-88
2	466	100.0	95	12	US-10-193-002-89
3	466	100.0	95	12	US-10-098-732A-35
4	466	100.0	100	12	US-10-084-843-115
5	466	100.0	100	12	US-10-193-002-110
6	466	100.0	100	16	US-10-080-170-639
7	466	100.0	358	9	US-09-287-849-8
8	466	100.0	358	12	US-10-359-460-8
9	466	100.0	802	9	US-09-287-849-10
10	466	100.0	802	12	US-10-084-843-214
11	466	100.0	802	12	US-10-084-843-351
12	466	100.0	802	12	US-10-193-002-209
13	466	100.0	802	12	US-10-193-002-346
14	466	100.0	802	12	US-10-359-460-10
15	462	99.1	100	15	US-10-140-045-5

16	392	84.1	80	12	US-10-084-843-117	Sequence 117, App
17	392	84.1	80	12	US-10-193-002-112	Sequence 112, App
18	217	46.6	49	15	US-10-140-045-6	Sequence 6, Appl
19	205	44.0	42	15	US-10-140-045-7	Sequence 7, Appl
20	173	37.1	100	15	US-10-140-045-28	Sequence 28, Appl
21	173	37.1	100	16	US-10-080-170-12	Sequence 12, Appl
22	137	29.4	28	15	US-10-140-045-12	Sequence 12, Appl
23	122	26.2	28	12	US-10-084-843-98	Sequence 98, Appl
24	121	26.0	27	12	US-10-084-843-96	Sequence 96, Appl
25	118	25.3	27	12	US-10-084-843-95	Sequence 95, Appl
26	117	25.3	27	12	US-10-084-843-97	Sequence 97, Appl
27	117	25.1	28	12	US-10-084-843-93	Sequence 93, Appl
28	100.5	21.6	1202	15	US-10-156-761-10195	Sequence 10195, A
29	100	21.5	20	15	US-10-140-045-8	Sequence 8, Appl
30	91	19.5	108	10	US-09-738-626-444	Sequence 4144, Ap
31	87.5	18.8	97	9	US-09-073-009-109	Sequence 109, App
32	87.5	18.8	97	9	US-09-073-009-143	Sequence 143, App
33	87.5	18.8	97	9	US-09-793-306-109	Sequence 109, App
34	87.5	18.8	97	9	US-09-793-306-143	Sequence 143, App
35	87.5	18.8	97	12	US-10-098-732A-24	Sequence 24, Appl
36	87.5	18.8	299	9	US-09-287-849-19	Sequence 19, Appl
37	87.5	18.8	299	12	US-10-359-460-19	Sequence 19, Appl
38	87.5	18.8	299	12	US-10-098-732A-47	Sequence 47, Appl
39	87.5	18.8	433	9	US-09-287-849-14	Sequence 14, Appl
40	87.5	18.8	433	12	US-10-359-460-14	Sequence 14, Appl
41	87.5	18.8	710	9	US-09-287-849-16	Sequence 16, Appl
42	87.5	18.8	710	12	US-10-359-460-16	Sequence 16, Appl
43	87.5	18.8	710	12	US-10-098-732A-49	Sequence 49, Appl
44	87.5	18.8	856	9	US-09-287-849-12	Sequence 12, Appl
45	87.5	18.8	856	12	US-10-359-460-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-10-084-843-88
; Sequence 88, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, David C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.

```
;
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-10-084-843-88
Query Match 100.0%; Score 466; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.9e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60

Qy 61 KOELDEISTNIRQAGVOYSRADBEQQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVOYSRADBEQQQALSSQMGF 95

RESULT 2
US-10-193-002-89
; Sequence 89, Application US/10193002
; Publication No. US20030135028A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
```

```
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 89:
US-10-193-002-89
Query Match 100.0%; Score 466; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.9e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60

Qy 61 KOELDEISTNIRQAGVOYSRADBEQQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVOYSRADBEQQQALSSQMGF 95

RESULT 3
US-10-098-732A-35
; Sequence 35, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Tb38-1 or 38-1 (MTb11)
US-10-098-732A-35
Query Match 100.0%; Score 466; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.9e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60

Qy 61 KOELDEISTNIRQAGVOYSRADBEQQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVOYSRADBEQQQALSSQMGF 95

RESULT 4
US-10-084-843-115
; Sequence 115, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
```


AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 115:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 115:

US-10-084-843-115

Query Match 100.0%; Score 466; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy. 1 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 60
Db 6 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 65

Qy 61 KOELDEISTNIRQAGVQYSRADDEQQOALLSSQMGF 95
Db 66 KOELDEISTNIRQAGVQYSRADDEQQOALLSSQMGF 100

RESULT 5

US-10-193-002-110
Sequence 110, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 110:

US-10-193-002-110

Query Match 100.0%; Score 466; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 60
Db 6 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 65

Qy 61 KOELDEISTNIRQAGVQYSRADDEQQOALLSSQMGF 95
Db 66 KOELDEISTNIRQAGVQYSRADDEQQOALLSSQMGF 100

RESULT 6

US-10-080-170-639
Sequence 639, Application US/10080170
Publication No. US20030129601A1
GENERAL INFORMATION:
APPLICANT: COLE, S.T.
TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
FILE REFERENCE: 03495.0218
CURRENT APPLICATION NUMBER: US/10/080,170
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/270,123
PRIOR FILING DATE: 2001-02-22
NUMBER OF SEQ ID NOS: 652
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 639
LENGTH: 100
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis

US-10-080-170-639

Query Match 100.0%; Score 466; DB 16; Length 100;
Best Local Similarity 100.0%; Pred. No. 2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 60
Db 6 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 65

Qy 61 KOELDEISTNIRQAGVQYSRADDEQQOALLSSQMGF 95

```

Db      66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
|||||
RESULT 7
US-09-287-849-8
; Sequence 8, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; CURRENT FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-09-287-849-8
Query Match 100.0%; Score 466; DB 9; Length 358;
Best Local Similarity 100.0%; Pred. No. 9.6e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TDAATLAQAGNFERISGDLKTQIDOVSTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 60
|||||
Db      264 TDAATLAQAGNFERISGDLKTQIDOVSTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 323
|||||
QY      61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
|||||
Db      324 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 358
|||||
RESULT 8
US-10-359-460-8
; Sequence 8, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT FILING DATE: 2003-02-05

```

```

; PRIOR APPLICATION NUMBER: US 09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TbH9-Tb38-1
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-10-359-460-8
Query Match 100.0%; Score 466; DB 12; Length 358;
Best Local Similarity 100.0%; Pred. No. 9.6e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TDAATLAQAGNFERISGDLKTQIDOVSTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 60
|||||
Db      264 TDAATLAQAGNFERISGDLKTQIDOVSTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 323
|||||
QY      61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
|||||
Db      324 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 358
|||||
RESULT 9
US-09-287-849-10
; Sequence 10, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```

OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-10

Query Match 100.0%; Score 466; DB 9; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.6e-42;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
DB 429 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 488
QY 61 KOELDEISTNIRQAGVQVSRADDEEQOQALSSQMGF 95
DB 489 KOELDEISTNIRQAGVQVSRADDEEQOQALSSQMGF 523

RESULT 10

US-10-084-843-214
Sequence 214, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 214:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 214:
US-10-084-843-214

Query Match 100.0%; Score 466; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.6e-42;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60

DB 429 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 488
QY 61 KOELDEISTNIRQAGVQVSRADDEEQOQALSSQMGF 95
DB 489 KOELDEISTNIRQAGVQVSRADDEEQOQALSSQMGF 523

RESULT 11

US-10-084-843-351
Sequence 351, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 351:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 351:
US-10-084-843-351

Query Match 100.0%; Score 466; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.6e-42;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
DB 429 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 488
QY 61 KOELDEISTNIRQAGVQVSRADDEEQOQALSSQMGF 95
DB 489 KOELDEISTNIRQAGVQVSRADDEEQOQALSSQMGF 523

RESULT 12

US-10-193-002-209
; Sequence 209, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Selsky, Yasir A.W.
Dillon, Davin C.
Campos-Meto, Antonia
Houghton, Raymond
Tvedick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA, 3001
ZIP: 98104-7092
COMPUTER: IBM PC compatible
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

CONSENT APPLICATION DATA
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>

APPLICATION NUMBER: US/09/072,596
 FILING DATE: 05-MAY-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.

REGISTRATION NUMBER: 31.392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
MATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 209:

US-10-193-002-209

Query Match	100.0%;	Score 466;	DB 12;	Length 802;
Best Local Similarity	100.0%;	Pred. No. 2.6e-42;		
Matches 95;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

1	TDAATLQAEAGNFERISGDKLTQIDQVESTAGSLQGWIRGAAGTAAQAAVWFQEAANKQ	60
	429	TDAATLQAEAGNFERISGDKLTQIDQVESTAGSLQGWIRGAAGTAAQAAVWFQEAANKQ
	61	KQELDEISTNIRQAGVOYSRADDEEQOALSSOMGF
	489	KQELDEISTNIRQAGVOYSRADDEEQOALSSOMGF

RESULT 13

US-10-193-002-346
; Sequence 346, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.

; CURRENT APPLICATION NUMBER: US/10/359,460
 ; CURRENT FILING DATE: 2003-02-05
 ; PRIOR APPLICATION NUMBER: US/09/287,849
 ; PRIOR FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 802
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
 US-10-359-460-10

Query Match 100.0%; Score 466; DB 12; Length 802;
 Best Local Similarity 100.0%; Pred. No. 2.6e-42;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 60
 Db 429 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 488
 QY 61 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
 Db 489 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 523

RESULT 15
 US-10-140-045-5
 ; Sequence 5, Application US/10140045
 ; Publication No. US20030092899A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GICQUEL, BRIGITTE
 ; APPLICANT: BERTHET, FRANCOIS-XAVIER
 ; APPLICANT: ANDERSEN, PETER
 ; APPLICANT: RASMUSSEN, PETER B
 ; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS VECTORS FOR THE SAME
 ; TITLE OF INVENTION: USING THE SAME
 ; FILE REFERENCE: 0660-0137-27X
 ; CURRENT APPLICATION NUMBER: US/10/140,045
 ; CURRENT FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: US/09/116,492A
 ; PRIOR FILING DATE: 1998-07-16
 ; PRIOR APPLICATION NUMBER: 60/252,631
 ; PRIOR FILING DATE: 1997-07-16
 ; NUMBER OF SEQ ID NOS: 39
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 100
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-10-140-045-5

Query Match 99.1%; Score 462; DB 15; Length 100;
 Best Local Similarity 98.9%; Pred. No. 5.5e-43;
 Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 60
 Db 6 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 65
 QY 61 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95

Db 66 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
 Search completed: November 21, 2003, 16:38:12
 Job time : 4.70522 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 / Search time 1.98186 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 TDAATLAQAGNFRISGDL.....VOYSGRADEQQQALSSQMGF 95

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76.*

1: Pirl.*

2: Pirl.*

3: Pirl.*

4: Pirl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	100	2	H70802
2	173	37.1	100	2	T10032
3	93.5	20.1	1186	2	T35661
4	90.5	19.4	410	2	B59103
5	90	19.3	750	2	T38435
6	87.5	18.8	97	2	T30836
7	86	18.5	103	2	B70600
8	86	18.5	2022	2	T43234
9	84.5	18.1	97	2	G70857
10	81.5	17.5	955	2	A47334
11	80	17.2	527	2	S33068
12	80	17.2	1940	2	A59287
13	79.5	17.1	245	2	T44704
14	79.5	17.1	247	2	G87116
15	79.5	17.1	1361	1	A61231
16	79.5	17.1	1999	1	S21801
17	79	17.0	508	1	K5SH12
18	78	16.7	1325	2	T42722
19	76	16.3	491	2	S05408
20	76	16.3	2094	2	S33124
21	75	16.1	163	2	P75559
22	74.5	16.0	80	2	D29674
23	74.5	16.0	80	2	A61873
24	74.5	16.0	316	2	JG6549
25	74.5	16.0	547	2	H82503
26	74	15.9	528	2	B75310
27	74	15.9	2101	2	A42184
28	73.5	15.8	433	2	D84335
29	73.5	15.8	440	1	S33479

ALIGNMENTS

RESULT 1

H70802

hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C/Accession: H70802

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, R.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: H70802

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-100 <COL>

A/Cross-references: GB:AL022120; GB:AL123456; NID:G3261558; PIDN:CAA17966.1; PID:G296022

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: RV3874

Query Match	100.0%;	Score 466;	DB 2;	Length 100;
Best Local Similarity	100.0%;	Pred. No. 9e-36;		
Matches	95;	Conservative	0;	Mismatches 0;
		Indels	0;	Gaps 0;
Qy	1	TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAQAAVVFQEAANKQ	60	
Db	6	TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAQAAVVFQEAANKQ	65	
Qy	61	KQELDEISTNIRQAGVOYSGRADEEQQALSSQMGF	95	
Db	66	KQELDEISTNIRQAGVOYSGRADEEQQALSSQMGF	100	

RESULT 2

T10032

hypothetical protein MLCB628.13c - Mycobacterium leprae

C/Species: Mycobacterium leprae

C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000

C/Accession: T10032

R/Higmeier, K.; Honore, N.; Woods, S.A.; Caudron, B.; Cole, S.T.

Mol. Microbiol. 7, 197-206, 1993

A/Title: Use of an ordered cosmid library to deduce the genomic organization of Mycobact

A/Reference number: Z16917; MUID:93188700; PMID:8446027

A/Accession: T10032

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-100 <BIG>

A/Cross-references: EMBL:Y14967; NID:G2370268; PIDN:CAA75210.1; PID:G2370280

Query Match 37.1%; Score 173; DB 2; Length 100;

Best Local Similarity 37.9%; Pred. No. 4.4e-09; Mismatches 20; Conservative 39; Indels 0; Gaps 0;
Matches 36; Conservative 39; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNPERISGLDKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
DB 6 TEAAITLQQAQPDQIAGSLGQERNFVDSIGQSFQNTWEGQAASAAALGALGRFDEAMQDQ 65
QY 61 KQELDEISTNIRQAGVOYSRADDEEQOALSSQMGP 95
DB 66 IQLESIVDKLNRSNGNYTKTDDEANQLLSKMF 100
RESULT 3
T35661
probable chromosome associated protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 02-Jun-2000
C:Accession: T35661
R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z21585
A:Accession: T35661
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1186 <MUR>
A:Cross-references: EMBL:AL034447; PIDN:CAA22420.1; GSPDB:GN00070; SCOEDB:SC7A1.21
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC7A1.21
C:Superfamily: chromosome segregation protein SMC1
Query Match 20.1%; Score 93.5; DB 2; Length 1186;
Best Local Similarity 33.0%; Pred. No. 1.1;
Matches 31; Conservative 15; Mismatches 41; Indels 7; Gaps 3;
QY 2 DAATLAQAGNPERISGLDKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANK 59
DB 691 ECAALVELGERRARRADREKSSVAQ---QLGLAGQARGAAGEAERSAAAARAEALDK 747
QY 60 KQELDEISTNIRQAGVOYSRADDEEQOALSSQM 93
DB 748 ALTEVELAB--RLVAENPVEEEDTAARDL 779
RESULT 4
B59103
hypothetical protein pX01-98 - Bacillus anthracis virulence plasmid pX01
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C:Accession: B59103
R:Okinkaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harbored
A:Reference number: A59091; MUID:59445483; PMID:10515943
A:Accession: B59103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32402.1; PID:g4894314
A:Experimental source: strain Sterne
C:Genetics:
A:Gene: pX01-98
A:Genome: plasmid
Query Match 19.4%; Score 90.5; DB 2; Length 410;
Best Local Similarity 29.5%; Pred. No. 0.67;
Matches 26; Conservative 14; Mismatches 27; Indels 21; Gaps 3;
QY 3 AATLAQAGNPERISGLDKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQK- 61
DB 16 AKTVNTRSSLEYTHQLYSQTEYIAS-----QMSGASSD-----RFYQMFNEAKP 61
QY 62 -----QELDEISTNIRQAGVOYSRADDE 83

DB 62 MMFNILQELDKIAVELERAANKFEADE 89
RESULT 5
T38435
coiled coil protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38435
R:McDougall, R.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z21793
A:Accession: T38435
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-750 <MCD>
A:Cross-references: EMBL:AL009227; PIDN:CAA15821.1; GSPDB:GN00066; SPDB:SPAC27D7.02c
A:Experimental source: strain 972h-; cosmid c27D7
C:Genetics:
A:Gene: SPDB:SPAC27D7.02c
A:Map position: 1
Query Match 19.3%; Score 90; DB 2; Length 750;
Best Local Similarity 28.2%; Pred. No. 1.4;
Matches 24; Conservative 18; Mismatches 23; Indels 20; Gaps 3;
QY 8 QEAG--NFERISGLDKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQKQELD 65
DB 550 KQAGENHYNLSSDYFIQIKLSLSS-----TNSQAECSYFQEKINELNSQID 597
QY 66 EISTNIRQAGVOYSRADDEEQOALS 90
DB 598 ELKLNKNEANKY-----QELAIS 616
RESULT 6
E70836
probable transcription regulator Rv0287 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70836
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70836
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-97 <COL>
A:Cross-references: GB:AL021930; GB:AL123456; NID:g3261524; PIDN:CAA17362.1; PID:e125247
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv0287
Query Match 18.8%; Score 87.5; DB 2; Length 97;
Best Local Similarity 34.9%; Pred. No. 0.26;
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;
QY 13 FERISGLDKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQKQELD EISTNIR 72
DB 18 FAKAGLNRHTIGQAQQAASQAQAFHQGESSAAQQAHAHFVAAAKVILLDDVAQNLG 77
QY 73 QAGVOYSRADDEEQOALSSQMGP 95
DB 78 EAAGTYVAAD---AAAASTVTGF 97
RESULT 7
E70600

hypothetical protein RV3805c - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: B70600
R/Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: B70600
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-103 <COL>
A/Cross-references: GB:Z94121; GB:AL123456; NID:g3261736; PIDN:CAB08096.1; PID:e312272;
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: RV3905c

Query Match 18.5%; Score 86; DB 2; Length 103;
Best Local Similarity 23.9%; Pred. No. 0.38;
Matches 21; Conservative 20; Mismatches 47; Indels 0; Gaps 0;

QY 2 DAATLAQAGNFERISGLKTKQIDOVSTAGSLQGWGAGTAQAQAVRFOEANKQK 61
DB 10 EPVAVQGGFAALDGAHGLAVQLAELDQVGMGLGWGASGAYSAGWELWHRGAGEVQ 69
QY 62 QELDEISTNIRQAGVYSRADEEQOALSOM 89
DB 70 LGLSLMAAAIAHAGAGYQHNETASQVL 97

RESULT 8

T43214
N/Title: protein - nematode (Onchocerca volvulus)
N/Alternate names: myosin-like antigen
C/Species: Onchocerca volvulus
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 20-Jun-2000
C/Accession: T43214; A44939; A54513; S27825
R/Rittertrapapab, S.; Ritchie, T.L.; Tuan, R.S.; Shepley, K.J.; Dinman, J.D.; Neubert, T.
Mol. Biochem. Parasitol. 69, 161-171, 1995
A/Title: Molecular cloning of a gene expressed during early embryonic development in Onchocerca
A/Reference number: Z22341; MUID:95287898; PMID:7770081
A/Accession: T43214
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-2022 <TRI>
A/Cross-references: EMBL:U12681; NID:G530824; PID:G530825; PIDN:AAA80009.1
A/Experimental source: specific host Homo sapiens
R/Erondou, N.E.; Donelson, J.E.
Mol. Biochem. Parasitol. 40, 213-224, 1990
A/Title: Characterization of a myosin-like antigen from Onchocerca volvulus.
A/Reference number: A44939; MUID:90301142; PMID:2194123
A/Accession: A44939
A/Molecule type: mRNA
A/Residues: 733-874, 'E', 876-916, 'S', 918-1038, 1040-1047, 'S', 1049-1283, 'E', 1285-1363 <ERO>
A/Cross-references: GB:M30398
A/Note: the sequence is revised in GenBank entry ONCANTML, release 115, (PIDN:AAA29413.1
R/Donelson, J.E.; Duke, S.O.B.; Moser, D.; Zeng, W.; Erondou, N.E.; Lucius, R.; Renz, A.
Mol. Biochem. Parasitol. 31, 241-250, 1988
A/Title: Construction of Onchocerca volvulus cDNA libraries and partial characterization
A/Reference number: A54513; MUID:89127417; PMID:2464764
A/Accession: A54513
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 733-866 <DON>
A/Cross-references: GB:J03095; NID:g159874; PIDN:AAA29412.1; PID:g159875
R/Ngizi, E.; Erondou, N.E.; Donelson, J.E.
submitted to the EMBL Data Library, April 1990
A/Description: Characterization of a myosin-like antigen from Onchocerca volvulus.
A/Reference number: S27825
A/Accession: S27825

A/Molecule type: mRNA
A/Residues: 733-874, 'E', 876-916, 'S', 918-1038, 1040-1047, 'S', 1049-1283, 'E', 1285-1347, 'E', 1
A/Cross-references: EMBL:M30398; NID:g159876; PIDN:AAA29413.1; PID:g159877
C/Genetics:
A/Gene: Ov1.
C/Keywords: leucine zipper
Query Match 18.5%; Score 86; DB 2; Length 2022;
Best Local Similarity 25.5%; Pred. No. 9.8;
Matches 25; Conservative 22; Mismatches 41; Indels 10; Gaps 2;
QY 6 LAQAGNFERISGLKTKQIDOVSTAGSLQGWGAGTAQAQAVRFOE 55
DB 1752 LDEKRTMENILHETALQREATLESSLNALERENKELHRNCAQLQOQIAQLENGNRLIQ 1811
QY 56 AANKQOELDEISTNIRQAGVYSRADEEQOALSOM 93
DB 1812 LTNKQREYDKFAQNWRTEKIQIERIENRSLKRSRI 1849

RESULT 9

G70857
Probable PE protein - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: G70857
R/Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: G70857
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-97 <COL>
A/Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CAA16105.1; PID:e123776
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: PE

Query Match 18.1%; Score 84.5; DB 2; Length 97;
Best Local Similarity 34.9%; Pred. No. 0.49;
Matches 29; Conservative 7; Mismatches 44; Indels 3; Gaps 1;

QY 13 FERISGLKTKQIDOVSTAGSLQGWGAGTAQAQAVRFOEANKQOELDEISTNIR 72
DB 18 FFAKAGIMRHTICQAEQQAMSAQAAPHQGSAAAFQGAHARFVAAAANKVNTLIDIAQANLG 77
QY 73 QAGVYSRADEEQOALSOMGF 95
DB 78 EAAGTYAAD---AAAASYTGF 97

RESULT 10

A47334
Lckin kinesin-related antigen - Leishmania chagasi (fragment)
C/Species: Leishmania chagasi
C/Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 02-Feb-2001
C/Accession: A47334
R/Burns Jr., J.M.; Shreffler, W.G.; Benson, D.R.; Ghalib, H.W.; Badaro, R.; Reed, S.G.
Proc. Natl. Acad. Sci. U.S.A. 90, 775-779, 1993
A/Title: Molecular characterization of a kinesin-related antigen of Leishmania chagasi
A/Reference number: A47334; MUID:93133867; PMID:8421715
A/Accession: A47334
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-955 <BUR>
A/Cross-references: GB:I07879; NID:g308884; PIDN:AAA29254.1; PID:g308885
A/Experimental source: MHOM/BJ/82/BA-2, C1
A/Note: sequence extracted from NCBI backbone (NCBI:122864, NCBI:122865)
C/Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C;Keywords: ATP; nucleotide binding; P-loop
F;13-398/Domain: kinesin motor domain homology <KMOT>
F;122-129/Region: nucleotide-binding motif A (P-loop)

Query Match 17.5%; Score 81.5; DB 2; Length 955;
Best Local Similarity 31.0%; Pred. No. 11;
Matches 31; Conservative 12; Mismatches 44; Indels 13; Gaps 2;

QY 4 ATLAQAGNFERISGLKTDQV-----ESTAGSLQGWGAGTAAGTAQAQAVVR 52
DB 858 ATLEQLRSEERAEALASOLETTAAKMSAEQDRESTRATLEQLRDSERAEALASQL 917

QY 53 FOEANKQKQELDEISTNIRQAGVQYSDRADEEQQALSSQ 92

DB 918 EATAAKSAEQDR--ENTRAALEQLRDSERAEALASQ 955

RESULT 11

S33068 myosin heavy chain - fluke (Schistosoma mansoni) (fragment)

N;Alternate names: surface antigen, 200K

C;Species: Schistosoma mansoni

C;Date: 22-Nov-1993 #sequence_revision 06-Sep-1996 #text_change 13-Feb-1998

C;Accession: S33068

R;Suisson, L.M.A.; Masterson, C.P.; Tom, T.D.; McNally, M.T.; Lowell, G.H.; Strand, M.

J. Immunol. 149, 3612-3620, 1992

A;Title: Induction of protective immunity in mice using a 62-kDa recombinant fragment of

A;Reference number: A46514; MUID:93056536; PMID:1431131

A;Accession: S33068

A;Molecule type: mRNA

A;Residues: 1-527 <SOI>

A;Cross-references: EMBL:X65591

A;Note: the authors translated the codon CAA for residue 346 as Lys

C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: ATP; surface antigen

Query Match 17.2%; Score 80; DB 2; Length 527;
Best Local Similarity 22.8%; Pred. No. 8;
Matches 21; Conservative 21; Mismatches 44; Indels 6; Gaps 2;

QY 2 DAATLAQAGNFERISGLKTDQVSTAGSLQGWGAGTAAGTAQAQAVVRFOEANKQK 61

DB 341 DLKATQETVDDLRRVRLDEQLRRKEAIGGLSGKFEDEQGLVAQ-----LQKIKELQ 395

QY 62 QELDEISTNIRQAGVQYSDRADEEQQALSSQ 93

DB 396 TRIQEELEDEAPRAARSKAKSRQK-LESEL 426

RESULT 12

A59287

myosin heavy chain - fluke (Schistosoma mansoni) (strain Brazilian LE)

C;Species: Schistosoma mansoni

C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000

C;Accession: A59287

R;Weston, D.S.; Schmitz, J.; Kemp, M.; Kunz, W.

Mol. Biochem. Parasitol. 58, 161-164, 1993

A;Title: Cloning and sequence characterization of a complete myosin heavy chain cDNA from

A;Reference number: A59287; MUID:93211444; PMID:8459827

A;Accession: A59287

A;Status: Preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-1940 <WES>

A;Cross-references: GB:L01634; PIDN:AAA29905.1

A;Experimental source: strain Brazilian LE

C;Genetics:

A;Gene: MYH

C;Superfamily: myosin heavy chain; myosin motor domain homology

F;82-752/Domain: myosin motor domain homology <MMO>

Query Match 17.2%; Score 80; DB 2; Length 1940;
Best Local Similarity 22.8%; Pred. No. 33;
Matches 21; Conservative 21; Mismatches 44; Indels 6; Gaps 2;

QY 2 DAATLAQAGNFERISGLKTDQVSTAGSLQGWGAGTAAGTAQAQAVVRFOEANKQK 61
DB 1044 DLKATQETVDDLRRVRLDEQLRRKEAIGGLSGKFEDEQGLVAQ-----LQKIKELQ 1098

QY 62 QELDEISTNIRQAGVQYSDRADEEQQALSSQ 93

DB 1099 TRIQEELEDEAPRAARSKAKSRQK-LESEL 1129

RESULT 13

T44704

hypothetical protein MLCB1243.13 [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 28-Jul-2000

C;Accession: T44704

R;Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1998

A;Reference number: Z22830

A;Accession: T44704

A;Status: preliminary; translated from GB/EMBL/DDBU

A;Molecule type: DNA

A;Residues: 1-245 <PAR>

A;Cross-references: EMBL:AL023635; PIDN:CAA19194.1

A;Experimental source: cosmid B1243

C;Genetics:

A;Note: MLCB1243.13

C;Superfamily: Mycobacterium leprae hypothetical protein MLCB1243.13

Query Match 17.1%; Score 79.5; DB 2; Length 245;

Best Local Similarity 27.0%; Pred. No. 3.9;

Matches 27; Conservative 20; Mismatches 42; Indels 11; Gaps 4;

QY 1 TDAATLAQEA-GNFERISGLKTDQVSTAGS-----LOGQWRGAGTAQAQAVVRFOE 55

DB 79 TESESLSHARAEADRIILSDAKSQVDRMASEAQHSERMLGDRBESIRIATVAKREYEA 138

QY 56 AANKQKQELDEISTNIRQAGVQYSDRADEE---QQQALSSQ 92

DB 139 SLNRAQSECDL---IENGNIISYEKAIQEGIKSQORLVSQ 175

RESULT 14

GS7116

conserved hypothetical protein MLI661 [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C;Accession: GS7116

R;Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho

R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sz

A;Title: Massive gene decay in the leprosy bacillus

A;Reference number: A86909; MUID:21128732; PMID:11234002

A;Accession: GS7116

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-247 <STO>

A;Cross-references: GB:AL450380; MUID:g13093432; PIDN:CAC30614.1; GSFDB:GN00147

C;Genetics:

A;Gene: MLI661

C;Superfamily: Mycobacterium leprae hypothetical protein MLCB1243.13

Query Match 17.1%; Score 79.5; DB 2; Length 247;

Best Local Similarity 27.0%; Pred. No. 3.9;

Matches 27; Conservative 20; Mismatches 42; Indels 11; Gaps 4;

QY 1 TDAATLAQEA-GNFERISGLKTDQVSTAGS-----LOGQWRGAGTAQAQAVVRFOE 55

DB 81 TESESLSHARAEADRIILSDAKSQVDRMASEAQHSERMLGDRBESIRIATVAKREYEA 140

QY 56 AANKQKQELDEISTNIRQAGVQYSDRADEE---QQQALSSQ 92

Search completed: November 21, 2003, 16:09:53
Job time : 2.98186 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 1.12018 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-10
Perfect score: 466
Sequence: 1 TDAATLAQAGNFERISGL.....VOYGRADEQQALSSQMGF 95

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	99	CF10_MYCTU	O69739 mycobacteri
2	173	37.1	99	CF10_MYCLE	O33084 mycobacteri
3	86	18.5	103	ES6D_MYCTU	O05440 mycobacteri
4	86	18.5	2022	ANT1_ONCV0	P21249 onchocerca
5	81.5	17.5	955	KINL_LEICH	P46865 leishmania
6	79.5	17.1	245	Y727_MYCLE	O9cbs6 mycobacteri
7	79.5	17.1	1960	MYH9_HUMAN	P35579 homo sapien
8	78	16.7	1325	G160_MOUSE	P35937 mus musculu
9	76	16.3	491	K2M2_SHEEP	P15241 ovis aries
10	76	16.3	2349	TPK_HUMAN	P12270 homo sapien
11	75.5	16.2	440	FENR_ANAVA	O44549 anabaena va
12	75.5	16.2	1961	MYH9_RAT	O62812 rattus norv
13	74.5	16.0	80	FENR_ANASP	P07124 anabaena sp
14	73.5	15.8	440	FENR_ANASO	P21890 anabaena sp
15	73.5	15.8	440	FENR_ANASP	P58558 anabaena sp
16	73.5	15.8	1679	Y109_YEAST	P40457 saccharomyc
17	73	15.7	1238	SBCR_RHOCA	O68032 rhodobacter
18	72.5	15.6	80	PXS1_MASLA	P11396 mastigoclad
19	72.5	15.6	1947	MYSC_CABEL	P12845 caenorhabdi
20	72	15.5	502	K2M3_SHEEP	P25691 ovis aries
21	71.5	15.3	292	BHC3_PSES1	P17297 pseudomonas
22	71.5	15.3	1172	LMB3_HUMAN	O13751 homo sapien
23	71	15.2	1938	MYSD_CABEL	P02567 caenorhabdi
24	70.5	15.1	78	PXS1_SYNEL	P50035 synechococc
25	70.5	15.1	189	APL3_MANGE	P13276 manduca sex
26	70.5	15.1	373	CK13_HUMAN	O02833 homo sapien
27	70.5	15.1	845	Y4PA_RHISN	P55439 rhizobium s
28	70.5	15.1	1959	MYH9_CHICK	P14105 gallus gall
29	70	15.0	186	APL3_GALME	P80703 gallaria me
30	70	15.0	512	MCFD_ENTAE	P21823 enterobacteri
31	69.5	14.9	242	GRPE_HALME	O9hnc2 halobacteri
32	69.5	14.9	457	MESE_LEUME	O10419 leuconostoc
33	69	14.8	232	YM29_MYCLE	O9cbs9 mycobacteri

34	69	14.8	384	1	K2CD_HUMAN	P48667 homo sapien
35	69	14.8	535	1	HTR1_HALNI	P33741 halobacteri
36	69	14.8	563	1	K2CE_HUMAN	P48668 homo sapien
37	68.5	14.7	245	1	Y727_MYCTU	O10973 mycobacteri
38	68.5	14.7	356	1	IRPA_SYNP7	P12608 synechococc
39	68.5	14.7	453	1	MSRE_BOVIN	P21258 bos taurus
40	68.5	14.7	576	1	PEXS_PICPA	P33292 pichia past
41	68.5	14.7	1742	1	MYSC_HUMAN	O9nqx4 homo sapien
42	68.5	14.7	2442	1	CEP2_HUMAN	O8bx73 homo sapien
43	68	14.6	327	1	SYFA_SALTY	O8xg13 salmonella
44	68	14.6	417	1	CZCC_ALCEU	P13509 alcaligenes
45	68	14.6	515	1	NACP_P13H4	P06159 human parai

ALIGNMENTS

RESULT 1
CF10_MYCTU
ID CF10_MYCTU STANDARD; PRT; 99 AA.
AC O69739;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 10 kDa culture filtrate antigen cfp10 (Secreted antigenic protein
MTSA-10).
DE CFP10 OR LHP OR MTSA10 OR RV3874 OR MT3988 OR MTV027.09.
GN Mycobacterium tuberculosis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
RC STRAIN=H37RV;
RX MEDLINE=99061212; PubMed=9846755;
RA Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,
RA Gicquel B.;
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
low-molecular-mass culture filtrate protein (CFP-10).";
RL Microbiology 144:3195-3203(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia P.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsey T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Emdinova M.D., Salzberg S.L.,
Deinzer A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bisshai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RP Singh B., Siddiqui Z., Singh S., Sharma P.;
RT "rv3874 (mtsA-10) gene of a clinical isolate of Mycobacterium
tuberculosis from India.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF004671; AAC83445.1; -
CC EMBL: AL022120; CAAL7966.1; -
CC EMBL: AE007190; AAK48356.1; -
CC EMBL: AF419854; AAL14999.1; -
CC DR PIR: H70802; H70802.
CC DR TIGR: MT3988; -
CC DR Tuberculin; Rv3874; -
CC KW Antigen; Complete proteome.
CC FT INIT MET 0
CC SQ SEQUENCE 99 AA; 10663 MW; EBCAE6A996C5489D CRC64;

Query Match 100.0%; Score 466; DB 1; Length 99;
Best Local Similarity 100.0%; Pred. No. 5.3e-36;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
DB 5 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 64

QY 61 KOELDEISTNIRQAGVOYSRADEEQQAALSSQMGF 95
DB 65 KOELDEISTNIRQAGVOYSRADEEQQAALSSQMGF 99

RESULT 2
CF10_MYCLE STANDARD; PRT; 99 AA.
ID CF10_MYCLE
AC O33084;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 10 kDa culture filtrate antigen cf10 homolog.
GN MLO050 OR MLCB628.13C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Eiglmeyer K., Garnier T., De Rossi E., Fsihi H., Cole S.T.,
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RT [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y14967; CAA75210.1; -
CC EMBL: ALS83917; CAC29958.1; -
CC DR PIR: T10032; T10032.
CC DR Leproma; MLO050; -
CC KW Complete proteome.
CC FT INIT MET 0
CC SQ SEQUENCE 99 AA; 10833 MW; 859B484F7E5E5A8A CRC64;

Query Match 37.1%; Score 173; DB 1; Length 99;
Best Local Similarity 37.9%; Pred. No. 2.3e-09;
Matches 36; Conservative 20; Mismatches 39; Indels 0; Gaps 0;

QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
DB 5 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 64

QY 61 KOELDEISTNIRQAGVOYSRADEEQQAALSSQMGF 95
DB 65 IRQLESIVDKLNRSGNYTKTDDDEANQLSSQMGF 99

RESULT 3
ES6D_MYCTU STANDARD; PRT; 103 AA.
ID ES6D_MYCTU
AC O05440;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ESAT-6 like protein 13.
GN RV3905C OR Mt4024 OR MCIY5F10.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL; Z94121; CAB08096.1; .....
DR EMBL; ABC07193; AAK48388.1; .....
DR PIR; B70600; B70600.
DR TIGR; MT4024; .....
DR Tuberculin; Bv3905c; .....
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 10460 MW; 3994E272A7BDF02 CRC64;

Query Match
Best Local Similarity 18.5%; Score 86; DB 1; Length 103;
Matches 21; Conservative 20; Mismatches 47; Indels 0; Gaps 0;

QY 2 DAATLAQEAAGNERISGDLTKTQIDQVESTAGSLQGWGAGTAQAQAQAAVVFQEAANKQK 61
Db 10 EFVAVQGFPAASDJDGAHHLAVGLAELDAQVQGLGWRGASGAYGSAMWELHWRGAGEVQ 69
QY 62 QELDIISTNIRQAGVQYSRADDEQQAL 89
Db 70 LGLSMLAAIAHAGAGYQHNETASQVL 97

RESULT 4
ANTI_ONCVO
ID ANTI_ONCVO STANDARD; PRT; 2022 AA.
AC P21249;
DT 01-MAY-1991 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Major antigen.
GN OVT1.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95287898; PubMed=7770081;
RA Triteaxaprab S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,
RA Neubert T.A., Scott A.L.;
RT "Molecular cloning of a gene expressed during early embryonic
development in Onchocerca volvulus."
RL Mol. Biochem. Parasitol. 69:161-171(1995).
RN [2]
RP SEQUENCE OF 733-866 FROM N.A.
RX MEDLINE=89127417; PubMed=2464764;
RA Donelson J.E., Duke B.O.L., Moser D., Zeng W., Erondou N.E.,
RA Lucius R., Renz A., Karam M., Flores G.Z.;
RT "Construction of Onchocerca volvulus cDNA libraries and partial
characterization of the cDNA for a major antigen."
RL Mol. Biochem. Parasitol. 31:241-250(1988).
CC -!- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U12681; AAA80009.1; -
CC EMBL; J03995; AAA29412.1; -
CC PIR; T43214; T43214.
KW Antigen; Coiled coil.
FT DOMAIN 74 120 COILED COIL (POTENTIAL).
FT DOMAIN 151 251 COILED COIL (POTENTIAL).
FT DOMAIN 327 384 COILED COIL (POTENTIAL).
FT DOMAIN 417 1879 COILED COIL (POTENTIAL).
SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;

Query Match
Best Local Similarity 18.5%; Score 86; DB 1; Length 2022;
Matches 25; Conservative 22; Mismatches 41; Indels 10; Gaps 2;

QY 6 LAQEAAGNERISGDLTKTQIDQVESTAGSLQGWGAGTAQAQAQAAVVFQEAANKQK 55
Db 1752 LDEKRTMENILHETALQREAISSNALERENKELHRNCAQLQQQLAQLENGNELIQ 1811
QY 56 AANKQKQELDEISTNIRQAGVQYSRADDEQQALSSOM 93
Db 1812 LTNKQREYDKFAQNWRTXIQIETIENRERSLKSR 1849

RESULT 5
KINL_LEICH
ID KINL_LEICH STANDARD; PRT; 955 AA.
AC P46865;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Kinesin-like protein K39 (Fragment).
GN KIN.
OS Leishmania chagasi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MEOM/BR/82 / Isolate BA-2;
RX MEDLINE=93133867; PubMed=8421715;
RA Burns J.W. Jr., Shreffler W.G., Benson D.R., Ghalib H.W., Badaro R.,
RA Reed S.G.;
RT "Molecular characterization of a kinesin-related antigen of
Leishmania chagasi that detects specific antibody in African and
American visceral leishmaniasis."
PL Proc. Natl. Acad. Sci. U.S.A. 90:775-779(1993).
CC -!- DEVELOPMENTAL STAGE: PREDOMINANT IN AMASTIGOTES.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; L07879; AAA29254.1; -
CC PIR; A47334; A47334.
CC HSP; P17119; 3KAR.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SMO0129; KISC; 1.
CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubules; ATP-binding; Coiled coil; Repeat.
FT DOMAIN 1 399 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 426 >955 COILED COIL (POTENTIAL).
FT NP_BIND 122 129 ATP (POTENTIAL).
FT DOMAIN 704 >955 7 X 39 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 704 742 1.
FT REPEAT 743 781 2.
FT REPEAT 782 820 3.
FT REPEAT 821 859 4.
FT REPEAT 860 898 5.
FT REPEAT 899 937 6.
FT REPEAT 938 >955 7 (PARTIAL).
FT NON_TER 955 955
SQ SEQUENCE 955 AA; 106168 MW; 8CA76815BE84C6E9 CRC64;

Query Match
Best Local Similarity 17.5%; Score 81.5; DB 1; Length 955;
Matches 31; Conservative 12; Mismatches 44; Indels 13; Gaps 2;

QY 4 ATLAQEAAGNERISGDLTKTQIDQVESTAGSLQGWGAGTAQAQAQAAVVFQEAANKQK 52

```

```

Db 858 ATLEQQKRESEERAAELASQLESTTAAKMSAEQDRESTRATLEQQQLRDSEERAAELASQL 917
QY 53 FQEAANKQKELDEISNIRQAGVQVSRADSEQQALSSQ 92
Db 918 EATRAAKSABQDR--ENTRAALEQQQLRDSEERAAELASQ 955

RESULT 6
YT27 MYCLE
ID YT27 MYCLE STANDARD; PRT; 245 AA.
AC Q9CB86; 069467;
DT 16-OCT-2001 (Rel. 40, Created)
DT 26-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein MLI1661.
GN MLI1661 OR MLCB1243.13.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seager K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV2927C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; AL023635; CAAL1914.1;
DR EMBL; AL883922; CAC30614.1; ALT_INIT.
DR PIR; T44704; T44704.
DR Leprosa; MLI1661; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 245 AA; 27087 MW; C984D9A5FA49697A CRC64;

Query Match 17.1%; Score 79.5; DB 1; Length 245;
Best Local Similarity 27.0%; Pred. No. 1.9;
Matches 27; Conservative 20; Mismatches 42; Indels 11; Gaps 4;

QY 1 TDAATLAQEA-GNPERISGLTKIQDQVESTAGS----LOGQWRGAAGTAAQAQAVRFOE 55
Db 79 TESESLSHAPAEADRLSDAKSQVDRMASEARQHSERMLGDAREESIRIATVAKREYEA 138
QY 56 AANKQKQELDEISNIRQAGVQVSRADSE---QQQALSSQ 92
Db 139 SLNPAQSECDRL----IENGNIYSVEKATQEGIKQQLRVLSQ 175

RESULT 7
MYH9_HUMAN
ID MYH9_HUMAN STANDARD; PRT; 1960 AA.
AC P35579; 060805;
DT 01-JUN-1994 (Rel. 29, Created)
DT 28-FEB-1998 (Rel. 37, Last sequence update)
DT 28-DEC-2003 (Rel. 41, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,

```

```

DE GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.P., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burill W.D., Burton J., Cardigan C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Gratham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashregui-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsay Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L., Swann R.M.,
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.B., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shiruya H., Simon M.I., Dunanski J.P., Peyrard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [2]
RP SEQUENCE OF 1-1337 FROM N.A.
RX MEDLINE=92003925; PubMed=1912569;
RA Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
RA Arnaout M.A., Clayton L.K., Tenen D.G.;
RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'
RT cDNA clones, characterization of the protein, chromosomal
RT localization, and upregulation during myeloid differentiation.";
RL Blood 78:1826-1833(1991).
RN [3]
RP SEQUENCE OF 1-715 FROM N.A.
RX MEDLINE=91316803; PubMed=1960190;
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA Gdula D., Adelstein R.S., Weir L.;
RT "Human nonmuscle myosin heavy chains are encoded by two genes located
RT on different chromosomes";
RL Circ. Res. 69:530-539(1991).
RN [4]
RP SEQUENCE OF 714-1960 FROM N.A.
RX MEDLINE=90138958; PubMed=1967836;
RA Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;

```


RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity
 RT through alternative polyadenylation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
 RN [5]
 RP VARIANT DFN17 HIS-705.
 RX MEDLINE=20489856; PubMed=11023810;
 RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
 RA Mahare A.N., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
 RA Magrini U., Bellotti S., Heath K.E., Babcock M., Glucksmann M.J.,
 RA Aliprandis E., Bizzaro N., Denick R.J., Martignetti J.A.;
 RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
 RT Sebastian syndromes.";
 RL Nat. Genet. 26:106-108(2000).
 RN [6]
 RP VARIANTS MHA/FTNS/SBS LYS-93; CYS-702; CYS-1165; HIS-1424 AND
 RP LYS-1841.
 RX MEDLINE=20428192; PubMed=10973259;
 RA Seri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
 RA Gloggeri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apollito M.,
 RA Iolascon A., Zelande L.L., Savio A., Balduini C.L., Nixis P.,
 RA Magrini U., Bellotti S., Heath K.E., Babcock M., Glucksmann M.J.,
 RA Aliprandis E., Bizzaro N., Denick R.J., Martignetti J.A.;
 RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
 RT Sebastian syndromes.";
 RL Nat. Genet. 26:103-105(2000).
 RN [7]
 RP VARIANTS MHA ILE-1155 AND LYS-1841.
 RX MEDLINE=20428193; PubMed=10973260;
 RA Kelley M.J., Jawien W., Ortel T.L., Korczak J.F.;
 RA "Mutation of MYH9 encoding non-muscle myosin heavy chain A, in
 RT May-Hegglin anomaly.";
 RL Nat. Genet. 26:106-108(2000).
 CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CC CAPPING.
 CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY
 CC (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF FECHTNER SYNDROME
 CC (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
 CC WITH ADDITIONAL ALPORT-LIKE CLINICAL FEATURES OF SENSORINEURAL
 CC DEAFNESS, CATARACTS AND NEPHRITIS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME
 CC (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
 CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA17) WHICH IS
 CC CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND
 CC COCHLEOACCULAR DEGENERATION.
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; 282215; CAB05105.1; -;
 CC EMBL; M81105; AAA59888.1; -;
 CC EMBL; M69180; AAA61765.1; -;
 CC EMBL; M31013; AAA36349.1; -;
 CC PIR; A61231; A61231.
 CC HSSP; P10587; 1BR2.
 CC Genew; HGNC:7579; MYH9.
 CC MIM; 160775; -;

DR MIM; 153640; -;
 DR MIM; 155100; -;
 DR MIM; 603622; -;
 DR MIM; 605249; -;
 DR GO; 0005860; C:non-muscle myosin; TAS.
 DR InterPro; IPR000048; IQ_region
 DR InterPro; IPR001689; myosin_head.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PRODOM; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
 KW Coiled coil; Alkylation; Multigene family; Disease mutation;
 KW Deafness.
 FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
 FT DOMAIN 779 808 IQ.
 FT COILED COIL (POTENTIAL).
 FT NP_BIND 837 1926 ATP (POTENTIAL).
 FT DOMAIN 174 181 ACTIN-BINDING.
 FT DOMAIN 654 676 ACTIN-BINDING.
 FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
 FT VARIANT 93 93 N -> K (IN MHA).
 FT VARIANT 702 702 R -> C (IN FTNS).
 FT VARIANT 705 705 R -> H (IN DENA17).
 FT VARIANT 1155 1155 T -> I (IN MHA).
 FT VARIANT 1165 1165 R -> C (IN SES).
 FT VARIANT 1424 1424 D -> H (IN FTNS).
 FT VARIANT 1841 1841 E -> K (IN MHA).
 FT CONFLICT 53 55 EAI -> RGH (IN REF. 3).
 FT CONFLICT 660 660 T -> S (IN REF. 3).
 FT CONFLICT 869 869 T -> M (IN REF. 4).
 FT CONFLICT 931 931 C -> Y (IN REF. 4).
 FT CONFLICT 1240 1241 KG -> GR (IN REF. 4).
 Query Match 17.1%; Score 79.5; DB 1; Length 1960;
 Best local similarity 21.7%; Pred. No. 18;
 Matches 34; Conservative 23; Mismatches 29; Indels 71; Gaps 5;
 QY 6 LAQAGNFERISGDK---TQDOV-----EST 30
 Db 1738 LEEQGNTELDRLKKNLQIDQINTLNLSHQAQKNAQQLERONKELVKLQEM 1797
 QY 31 AGSLQGWGAGTAQAQAVVRQE-----AANKQ----- 60
 Db 1798 EGVKSKYK-ASITALEAKIAQLDNETKQAACKQVRTEKKIKDVLQVDDRR 1856
 QY 61 -----KQELDEITNIRQAGVYSRADEFCQQLSSQ 92
 Db 1857 NAEQYKQADKASTRLKQLKQLEEAEEEAQANASR 1893
 RESULT 8
 G160_MOUSE
 ID_G160_MOUSE STANDARD; PRT; 1325 AA.
 AC P55937;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Golgin-160 (Male-enhanced antigen-2) (MEA-2).
GN GOLGA3 OR MEA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Testis;
RX MEDLINE=97217683; PubMed=9063644;
RA Kondo M., Sutou S.;
RT "Cloning and molecular characterization of cDNA encoding a mouse
RT male-enhanced antigen-2 (Mea-2): a putative family of the Golgi
RT autoantigen."
RL DNA Seq. 7:71-82 (1997).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN SPERMATOGENESIS AND/OR
CC TESTIS DEVELOPMENT. PROBABLY IDENTICAL WITH THE SEROLOGICALLY
CC DETECTABLE MALE ANTIGEN (SDM).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. TRANSCRIPTS CAN BE
CC FOUND IN SPERMATIDS DURING SPERMATOGENESIS. NO EXPRESSION IN
CC LEYDIG CELLS, SPERMATOGENIA, OR SPERMATOCYTES.
CC -!- SIMILARITY: HIGH, TO HUMAN GOLGIN-160.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-19 OR MET-30 IS THE
CC INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D78270; BAA19612.1; --
DR PIR; T42722; T42722.
DR MGD; MGI:196958; Golga3.
DR GO; GO:0005793; C:ER-Golgi intermediate compartment; IDA.
DR GO; GO:0000139; C:Golgi membrane; IDA.
DR GO; GO:0005515; F:protein binding activity; IPI.
DR GO; GO:0005515; F:protein binding activity; IPI.
KW Spermatogenesis; Developmental protein.
FT DOMAIN 201 204 POLY-ALA.
SQ SEQUENCE 1325 AA; 149880 MW; 3230636962C687B0 CRC64;
Query Match 16.7%; Score 78; DB 1; Length 1325;
Best Local Similarity 25.5%; Pred. No. 16;
Matches 28; Conservative 23; Mismatches 33; Indels 26; Gaps 4;
QY 4 ATLAQAGNPERISGLDKTQIDOVSTAG-----SLOQWRGAAG 43
DB 201 AARACHQDQNGEVRSDICSVSMESLSAEPQDELLOILKDKRLEQVEALSJ 260
QY 44 TAAQAAVVRFOEANKQKQELDEISNIRQAGVOYGRADDEEQQALSSQM 93
DB 261 EASQA----LQKAEALQAO-LAALSTR-L-QAQVEHSHSQKQDLSLSEV 304
RESULT 9
K2M2 SHEEP STANDARD; PRT; 491 AA.
AC P15241;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II microfibrillar, component 7C.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE
RX MEDLINE=90026244; PubMed=2803231;
RA Sparrow L.G., Robinson C.P., McMahon D.T.W., Rubira M.R.;

RT "The amino acid sequence of component 7c, a type II intermediate-
RT filament protein from wool."
RL Biochem. J. 261:1015-1022 (1989).
CC -!- FUNCTION: WOOL MICROFIBRILLAR KERATIN.
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.
CC -!- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,
CC 7A, 7B, AND 7C).
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT MOD RES 1 109
FT DOMAIN 1 109
FT DOMAIN 110 416
FT DOMAIN 417 491
FT DOMAIN 110 144
FT DOMAIN 145 154
FT DOMAIN 155 255
FT DOMAIN 256 272
FT DOMAIN 273 416
FT DOMAIN 74 74
FT VARIANT 80 80
FT VARIANT 144 144
FT VARIANT 232 232
FT VARIANT 276 276
FT VARIANT 284 284
FT UNSURE 1 2
SQ SEQUENCE 491 AA; 53681 MW; A801771FE3831ABE CRC64;
Query Match 16.3%; Score 76; DB 1; Length 491;
Best Local Similarity 22.9%; Pred. No. 8.5;
Matches 19; Conservative 23; Mismatches 39; Indels 2; Gaps 1;
QY 12 NFERISGLDKTQIDOVSTAGSLQGWGRGAAGTAAQAAVVRFOEANKQKQELDEISN 71
DB 273 NDCIVAEIKAYQYDDIASRRAEAESEWYRSKCEIKATVIRHGETLRTTBELNELNVI 332
QY 72 RQ--AGVOYSRADDEEQQALSSQ 92
DB 333 QRLTAIVENAKQNSKLEAAVTQ 355
RESULT 10
TPR_HUMAN STANDARD; PRT; 2349 AA.
ID TPR_HUMAN
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain."
RN Oncogene 7:2329-2333 (1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.


```

DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN SEQUENCE FROM N.A.
RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC CAPPING.
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31463; AAA74950.1; --
DR HSPB; P10587; 13E2
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; Myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSc; 1.
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS50096; IQ; 1.
DR Myosin; ATP-binding; G-actin-binding; Actin-binding;
DR Coiled coil; Alkylation; Multigene family.
DR DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808
FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
FT MOD_RES 594 594 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1961 AA; 226336 MW; 9B9876D9681F819E CRC64;

Query Match 16.2%; Score 75.5; DB 1; Length 1961;
Best Local Similarity 22.8%; Pred. No. 41;
Matches 26; Conservative 26; Mismatches 23; Indels 39; Gaps 4;

Oy 14 ERISGDKLTQIDQVESTAGSLQGWRCAGTAQAQAVVRFQ-----AANKQ--- 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1785 ERQNKELKAKLQEMES---AVKYSK-ASIAALEAKIAQLQLEBQDNETKERRQAASQVRR 1840
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Oy 61 -----KQELDEISTNIRQAGVQYVSRADEEQQALSSQ 92
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1841 AEKKLKVLLQVEDERNARQFKDQADKASTRLKQLKQLEEAEEAQAANASR 1894
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 13
PYS1 ANASP
ID PYS1 ANASP
AC P07124;

```

```

DT 01-APR-1998 (Rel. 07, Created)
DT 01-APR-1998 (Rel. 07, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phycobilisome 8.9 kDa linker polypeptide, phycocyanin-associated, rod
DE (L-8.9/R) (Rod capping linker protein).
GN CPD OR ASR0531
OS Anabaena sp. (Strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN SEQUENCE FROM N.A.
RA Belknap W.R., Haselkorn R.;
RX MEDLINE=87246520; PubMed=3109890;
RT "Cloning and light regulation of expression of the phycocyanin operon
RT of the cyanobacterium Anabaena.";
RL EMBO J. 6:871-884(1987).
RN SEQUENCE FROM N.A.
RX MEDLINE=21134097; PubMed=11237320;
RA Cai Y.A., Murphy J.T., Wedemayer G.J., Glazer A.N.;
RT "Recombinant phycobiliproteins. Recombinant C-phycocyanins equipped
RT with affinity tags, oligomerization, and biospecific recognition
RT domains.";
RL Anal. Biochem. 290:186-204 (2001).
RN SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -1- SIMILARITY: TO OTHER PHYCOBILISOME LINKER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05239; CAA28865.1; --
DR EMBL; AF178757; AAG09319.1; --
DR EMBL; AP003592; BAB72489.1; --
DR FIR; AB1873; AB1873.
DR InterPro; IPR001685; CpcD-like.
DR Pfam; PF01383; CpcD; 1.
DR ProDom; PD002828; CpcD-like C; 1.
KW Phycobilisome; Photosynthesis; Complete proteome.
SQ SEQUENCE 80 AA; 8895 MW; 52DF2D7DEF544D5 CRC64;

Query Match 16.0%; Score 74.5; DB 1; Length 80;
Best Local Similarity 35.6%; Pred. No. 1.7;
Matches 21; Conservative 10; Mismatches 21; Indels 7; Gaps 2;

Oy 36 GQWRCAAGTAQAQAA--VVFQFAANKQKQELDEISTNIRQAG-----VOXSRADEEQQ 87
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 3 QQTILGAGSVSSASRFRVEVGLRQSSSETDKNKYINRSGSVFITVPYSRMNEEYOR 61
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 14
FENR ANASO
ID FENR ANASO STANDARD; PRT; 440 AA.
AC P21890;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ferredoxin-NADP reductase (EC 1.18.1.2) (FNR).
GN PETH.

```

OS Anabaena sp. (strain PCC 7119).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 ON NCBI_TaxID=1169;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93344523; PubMed=8343609;
 RA Fillat M.F., Flores E., Gomez-Moreno C.;
 RT "Homology of the N-terminal domain of the peth gene product from
 RT Anabaena sp. PCC 7119 to the CpcD phycobilisome linker polypeptide.";
 RL Plant Mol. Biol. 22:725-729(1993).
 RN [2]
 RP SEQUENCE OF 137-440 FROM N.A.
 RX MEDLINE=91088322; PubMed=2124680;
 RA Fillat M.F., Bakker H.A.C., Welsbeek P.J.;
 RT "Sequence of the ferredoxin-NADP(+) reductase gene from Anabaena PCC
 RT 7119.";
 RL Nucleic Acids Res. 18:7161-7161(1990).
 RN [3]
 RP SEQUENCE OF 152-193.
 RX STRAIN=1403.46;
 RA MEDLINE=8132819; PubMed=3124746;
 RA Sancho J., Peleato M.L., Gomez-Moreno C., Edmondson D.E.;
 RT "Purification and properties of ferredoxin-NADP+ oxidoreductase from
 RT the nitrogen-fixing cyanobacteria Anabaena variabilis.";
 RL Arch. Biochem. Biophys. 260:200-207(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 137-440.
 RX MEDLINE=97045988; PubMed=8890910;
 RA Serre L., Vellieux F.M.D., Medina M., Gomez-Moreno C.,
 RA Fontecilla-Camps J.C., Frey M.;
 RT "X-ray structure of the ferredoxin:NADP+ reductase from the
 RT cyanobacterium Anabaena PCC 7119 at 1.8-A resolution, and
 RT crystallographic studies of NADP+ binding at 2.25-A resolution.";
 RL J. Mol. Biol. 263:20-39(1996).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 146-440.
 RX MEDLINE=99119196; PubMed=9922134;
 RA Martinez-Julvez M., Hermoso J., Hurley J.K., Mayoral T.,
 RA Sanz-Aparicio J., Tollin G., Gomez-Moreno C., Medina M.;
 RT "Role of Arg100 and Arg264 from Anabaena PCC 7119 ferredoxin-NADP+
 RT reductase for optimal NADP+ binding and electron transfer.";
 RL Biochemistry 37:17680-17691(1998).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 137-440.
 RX MEDLINE=20114461; PubMed=10651039;
 RA Mayoral T., Medina M., Sanz-Aparicio J., Gomez-Moreno C.,
 RA Hermoso J.A.;
 RT "Structural basis of the catalytic role of Glu301 in Anabaena PCC
 RT 7119 ferredoxin-NADP+ reductase revealed by x-ray crystallography.";
 RL Proteins 38:60-69(2000).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.38 ANGSTROMS).
 RX MEDLINE=20508220; PubMed=11053838;
 RA Morales R., Kachalova G., Vellieux F., Charon M.-H., Frey M.;
 RT "Crystallographic studies of the interaction between the
 RT ferredoxin-NADP+ reductase and ferredoxin from the cyanobacterium
 RT Anabaena: looking for the elusive ferredoxin molecule.";
 RL Acta Crystallogr. D 56:1408-1412(2000).
 CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized
 CC ferredoxin + NADPH.
 CC -!- COFACTOR: FAD.
 CC -!- SUBCELLULAR LOCATION: May be bound to the thylakoid membrane or
 CC anchored to the thylakoid-bound phycobilisomes.
 CC -!- SIMILARITY: WITH OTHER SPECIES FNR.
 CC -!- SIMILARITY: Contains 1 cpcd-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X72394; CAA51088.1; -;
 DR EMBL; X54039; CAA37973.1; -;
 DR PDB; 1QUE; 15-MAY-97.
 DR PDB; 1QUF; 17-SEP-97.
 DR PDB; 1BJK; 18-NOV-98.
 DR PDB; 1B28; 15-DEC-99.
 DR PDB; 1EWY; 07-FEB-01.
 DR PDB; 1BQE; 27-FEB-02.
 DR PDB; 1E62; 14-JUN-01.
 DR PDB; 1E63; 14-JUN-01.
 DR PDB; 1E64; 14-JUN-01.
 DR PDB; 1GJR; 27-JUN-02.
 DR PDB; 1GO2; 17-OCT-02.
 DR PDB; 1GR1; 24-OCT-02.
 DR PDB; 1H85; 28-NOV-01.
 DR PDB; 1QGH; 27-FEB-02.
 DR PDB; 1QHO; 27-FEB-02.
 DR InterPro; IPR001685; CpcD-like.
 DR InterPro; IPR001834; Cyt B5 reductase.
 DR InterPro; IPR001709; FPN_Cyt reductase.
 DR InterPro; IPR001433; Oxred_FAD/NAD(P).
 DR Pfam; PF01383; CpcD; 1.
 DR Pfam; PF00175; NAD_binding_1; 1.
 DR PRINTS; PR00406; CYTB5REDTASE.
 DR PRINTS; PR00371; FPNCR.
 DR ProDom; PD002828; CpcD-like C; 1.
 KW Oxidoreductase; Flavoprotein; NADP; FAD; Thylakoid; Membrane;
 KW Phycobilisome; 3D-structure. CPCD-LIKE.
 FT DOMAIN 1 80
 FT NP_BIND 288 306
 FT CONFLICT 180 180
 FT STRAND 152 153
 FT TURN 154 155
 FT STRAND 156 156
 FT STRAND 158 167
 FT TURN 171 172
 FT STRAND 177 183
 FT TURN 185 186
 FT TURN 193 194
 FT STRAND 196 200
 FT STRAND 204 204
 FT TURN 206 207
 FT STRAND 210 210
 FT STRAND 214 218
 FT TURN 222 227
 FT STRAND 231 237
 FT STRAND 240 241
 FT STRAND 250 251
 FT HELIX 253 259
 FT TURN 260 260
 FT TURN 263 264
 FT STRAND 266 273
 FT TURN 283 284
 FT STRAND 286 292
 FT HELIX 293 295
 FT HELIX 296 307
 FT HELIX 309 314
 FT TURN 316 317
 FT STRAND 323 330
 FT HELIX 333 335
 FT TURN 337 338
 FT HELIX 339 348
 FT TURN 350 352
 FT STRAND 353 359
 FT TURN 360 362
 FT STRAND 364 364
 FT TURN 366 367
 FT STRAND 370 370
 FT HELIX 373 379
 FT TURN 380 380
 FT HELIX 381 388
 FT TURN 389 389

FT TURN 391 392
FT STRAND 393 399
FT TURN 401 402
FT HELIX 404 416
FT TURN 417 419
FT HELIX 422 431
FT TURN 432 433
FT STRAND 435 440
SQ SEQUENCE 440 AA; 48865 MW; 8E1F61D0F09338B6 CRC64;
Query Match 15.8%; Score 73.5; DB 1; Length 440;
Best Local Similarity 30.0%; Pred. NO. 13;
Matches 18; Conservative 15; Mismatches 22; Indels 5; Gaps 1;
QY 33 SLOGQWRGAAGTAQAQAAVRFQEAANKQKQELDEISTNIRQAG-----VOYSRADEEQQ 87
DB 2 SNQAFDGAANVESGRVVFYEVVGMKQNEETDQTNPIRKSGSVFIRVPYINRMQEMQR 61

Search completed: November 21, 2003, 16:04:18
Job time : 2.12018 secs

QY 33 SLOGQWRGAAGTAQAQAAVRFQEAANKQKQELDEISTNIRQAG-----VOYSRADEEQQ 87
DB 2 SNQAFDGAANVESGRVVFYEVVGMKQNEETDQTNPIRKSGSVFIRVPYINRMQEMQR 61

RESULT 15

FENR ANASP
ID FENR ANASP STANDARD; PRT; 440 AA.
AC P8658; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ferredoxin--NADP reductase (EC 1.18.1.2) (FNR).
GN PETH OR ALL4121.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001)
CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized
CC ferredoxin + NADPH.
CC -!- COFACTOR: FAD.
CC -!- SUBCELLULAR LOCATION: May be bound to the thylakoid membrane or
CC anchored to the thylakoid-bound phycobilisomes.
CC -!- SIMILARITY: WITH OTHER SPECIES FNR.
CC -!- SIMILARITY: Contains 1 cpcd-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003595; BAB75820.1; -
CC PIR; AB2321; AB2321.
CC HSP; P21890; 1B2R.
DR InterPro; IPR001685; CpcD-like.
DR InterPro; IPR001834; Cyt B5 reductase.
DR InterPro; IPR001709; FEN_cyt reductse.
DR InterPro; IPR001433; Oxred_FAD/NAD(P).
DR Pfam; PF01383; CpcD; 1.
DR Pfam; PF00175; NAD_binding_1; 1.
DR PRINTS; PR00406; CYTB5RDTASE.
DR PRINTS; PR00371; FPNCR.
DR ProDom; PD002828; CpcD-like C; 1.
DR Oxidoreductase; Flavoprotein; NADP; FAD; Thylakoid; Membrane;
KW Phycobilisome; Complete proteome.
FT DOMAIN 1 80 CPD-LIKE.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 4.6746 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 TDAATLAQEA G N F E R I S G D L VQY S R A D E E Q Q A L S S Q M G F 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*

```

1:  sp.archaea.*
2:  sp.bacteria.*
3:  sp.fungi.*
4:  sp.human.*
5:  sp.invertebrate.*
6:  sp.mammal.*
7:  sp_mhc.*
8:  sp_ordanelle.*
9:  sp.phage.*
10: sp.plant.*
11: sp.rodent.*
12: sp.virus.*
13: sp_vetbrate.*
14: sp_unclassified.*
15: sp.virus.*
16: sp.bacteriap.*
17: sp.archeap.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	114	24.5	106	16	Q8F324	corynebacte
2	93.5	20.1	1186	16	Q2B02	streptomyce
3	91	19.5	96	16	Q8G78	blifidomyce
4	91	19.5	108	16	Q3NST8	
5	90.5	19.4	410	2	Q9X367	corynebacte
6	90	19.3	750	3	Q42657	bacillus an
7	87.5	18.8	97	16	Q53692	schizosacch
8	84.5	18.1	97	16	Q53267	mycobacteri
9	84.5	18.1	102	16	Q8VJ87	mycobacteri
10	84	18.0	1200	11	Q921B9	mus musculu
11	83.5	17.9	188	5	Q77248	spodoptera
12	81	17.4	914	16	Q8ED18	shewanella
13	80.5	17.3	707	2	Q9F5N1	rhizobium m
14	80	17.2	528	5	Q26589	schistosoma
15	80	17.1	1940	5	Q02456	schistosoma
16	79.5	17.1	409	2	Q8VQM8	prevotella

17	79.5	17.1	474	2	Q8GI70	Q8GI70 rhodopseudo
18	79.5	17.1	1999	11	Q63731	Q63731 rattus norv
19	79	17.0	508	6	Q28582	Q28582 oviss aries
20	78	16.7	178	5	Q02415	Q02415 agrius conv
21	78	16.7	1447	11	Q9QYT2	Q9QYT2 mus musculus
22	78	16.7	1487	11	Q9QYT3	Q9QYT3 mus musculus
23	77.5	16.6	790	2	Q8ZQF5	Q8Zdf5 salmonella
24	77	16.5	507	4	P78386	P78386 homo sapien
25	77	16.5	507	4	Q9NSB1	Q9nsb1 homo sapien
26	77	16.5	801	16	Q8P3Q3	Q8p3q3 xanthomonas
27	76.5	16.4	459	16	Q8G7M4	Q8g7m4 bifidobacte
28	76.5	16.4	623	16	Q8D3R4	Q8d3r4 vibrio vuln
29	76.5	16.4	698	16	Q92BZ4	Q92br4 rhizobium m
30	76.5	16.4	1186	3	Q9N0J2	Q9njj2 aspergillus
31	76	16.3	333	2	Q9X7H6	Q9x7h6 paracoccus
32	76	16.3	401	16	Q8PLD30	Q8pl30 xanthomonas
33	76	16.3	671	13	Q9YHD4	Q9yhd4 rana catesb
34	76	16.3	1390	4	Q8WZ43	Q8wz43 homo sapien
35	76	16.3	1498	4	Q8TDA9	Q8tda9 homo sapien
36	76	16.3	1530	4	Q83241	Q83241 homo sapien
37	75.5	16.2	529	11	Q9QYJ3	Q9qyj3 mus musculus
38	75.5	16.2	622	16	Q8D3R4	Q8d3r4 vibrio vuln
39	75.5	16.2	1960	11	Q8VDD5	Q8vdd5 mus musculus
40	75.5	16.2	2756	10	Q9LJ60	Q9lj60 arabidopsis
41	75	16.1	163	16	Q9R346	Q9ry46 deinococcus
42	75	16.1	507	11	Q9D7M4	Q9d7m4 mus musculus
43	75	16.1	507	11	Q9Z2T6	Q9z2t6 mus musculus
44	75	16.1	507	11	Q8CEB3	Q8ceb3 mus musculus
45	75	16.1	1793	4	Q8TEP3	Q8tep3 homo sapien

ALIGNMENTS

RESULT 1

RECEIVED
Q8FS24

ID	Q8FS24	PRELIMINARY;	PRT;	106 AA.
AC	Q8FS24;			
DT	01-MAR-2003	(T-EMBLrel. 23, Created)		
DT	01-MAR-2003	(T-EMBLrel. 23, Last sequence update)		
DT	01-MAR-2003	(T-EMBLrel. 23, Last annotation update)		
DE	Hypothetical protein.			
GN	CE0582.			
OS	Corynebacterium efficiens.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriaceae; Corynebacterium.			
OX	NCBI_taxid=152794;			
PN	[1]			

RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawababayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.,
RT "The entire genomic sequence of *Corynebacterium efficiens* YS-314.";
RT RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005215; BAC17392.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 106 AA; 11382 MW; 73AF6CDBEA7838A6 CRC64;

Query Match 24.5%; Score 114; DB 16; Length 106;
Best Local Similarity 24.5%; Pred.No. 0.0051;
Matches 23; Conservative 25; Mismatches 46; Indels 0; Gaps 0;

RESULT 2

```

Q9ZBQ2          PRELIMINARY;          PRT; 1186 AA.
ID Q9ZBQ2
AC Q9ZBQ2
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Purative chromosome associated protein.
GN SC0577 OR SC7A1.21.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapeite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.; and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Chater K.F., Harris D.E., Cordenio-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris H.M., Denapeite D., Eichner A., Cullum J.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939124; CAA22420.1; -.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR InterPro; IPR005829; Sug_transporter.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR Complete proteome.
SQ SEQUENCE 1186 AA; 128723 MW; CB11027815373E99 CRC64;

Query Match 20.1%; Score 93.5; DB 16; Length 1186;
Best Local Similarity 33.0%; Pred. No. 5.1;
Matches 31; Conservative 15; Mismatches 41; Indels 7; Gaps 3;

QY 2 DAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGAGTA--AQAAVVFQEAANK 59
Db 691 ECAALVELGERRRAAREKSSVAQ---QLGRLAGQARGAAGERSAALAAERQALDK 747
QY 60 QKQELDEISTNIRQAGVQYSRADEEQQALSSQM 93
Db 748 ALTEVEELAE--RLVAEEMPEVEPTAARDRL 779

RESULT 3
Q8G878          PRELIMINARY;          PRT; 96 AA.
ID Q8G878
AC Q8G878
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN BL0003.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Fessi G., Zwanen M.-C., Desliere F., Bork P., Delley M.,
RA Fridmore R.D., Arigoni F.;
RL "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
DR EMBL; AE014614; AAN23870.1; -.
RW Hypothetical protein; Complete proteome
SQ SEQUENCE 96 AA; 10431 MW; 7D02B38C6B1F61C CRC64;

Query Match 19.5%; Score 91; DB 16; Length 96;
Best Local Similarity 27.3%; Pred. No. 0.53;
Matches 24; Conservative 17; Mismatches 37; Indels 10; Gaps 2;

QY 14 ERI---SDLLTQIDQVESTAG-----SLQGWGAGTAQAQAAVVFQEAANKQKE 63
Db 9 ERIQSSAAVATISQIRQAVGVMTNLNALQDARWGSAAQTFTAVAEQWRAAQQQMEAS 68
QY 64 LDEISTNIRQAGVQYSRADEEQQALSS 91
Db 69 LESIQRSITQASTVYADAEIQASRLFAS 96

RESULT 4
Q8NST8          PRELIMINARY;          PRT; 108 AA.
ID Q8NST8
AC Q8NST8
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein Cgl0579.
GN CGL0579.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005275; BAB97972.1; -.
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 108 AA; 11644 MW; EUB3AFA03AE9D452 CRC64;

Query Match 19.5%; Score 91; DB 16; Length 108;
Best Local Similarity 19.8%; Pred. No. 0.6;
Matches 18; Conservative 26; Mismatches 47; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGAGTAQAQAAVVFQEAANKQ 60
Db 7 TESDVMLATASQVDDINDQVQELSLRGVVDVRSWAGQAVSFSLMRWNSSARQL 66
QY 61 KOELDEISTNIRQAGVQYSRADEEQQALSS 91
Db 67 QEALASISDNIRHNARSFENTEADNSQAFNA 97

```



```

[2]
RN
RC SEQUENCE FROM N.A.
RP STRAIN=972h-;
RA Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL009227; CAA15821.1; -
DR GeneDB Spombe; SPAC27D7.02C; -
DR InterPro; IPR000237; GRIP_domain.
DR Pfam; PF01465; GRIP; 1.
SQ SEQUENCE 750 AA; 87265 MW; A469AD95C5787042 CRC64;

Query Match 19.3%; Score 90; DB 3; Length 750;
Best Local Similarity 28.2%; Pred. No. 6.3;
Matches 24; Conservative 18; Mismatches 23; Indels 20; Gaps 3

QY 8 QEAG--NFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRQEAANKQKQQLD 65
DB 550 KQAGENHYNLSGSDYETQIKSLESL-----TNSQAEVSPQEKINELNSQID 597
QY 66 EISTNRQAGVOYSGRADEEQOALS 90
DB 598 ELKLNKNEANKY-----QELALS 616

RESULT 7
O53692
ID O53692 PRELIMINARY; PRT; 97 AA.
AC O53692;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 06, Last annotation update)
DE Hypothetical 9.8 kDa protein (PE family protein).
GN RV0287 OR MT035.15 OR MT0300.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL021930; CAA17362.1; -
DR ENBL; AS008937; AAK44524.1; -
DR TIGR; MT0300; -
DR Tuberculist; RV0287; -
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 97 AA; 9778 MW; 927527DA610A1637 CRC64;

Query Match 18.8%; Score 87.5; DB 16; Length 97;
Best Local Similarity 34.9%; Pred. No. 1.1;
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1

RN
RC SEQUENCE FROM N.A.
RP STRAIN=972h-;
RA Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL009227; CAA15821.1; -
DR GeneDB Spombe; SPAC27D7.02C; -
DR InterPro; IPR000237; GRIP_domain.
DR Pfam; PF01465; GRIP; 1.
SQ SEQUENCE 750 AA; 87265 MW; A469AD95C5787042 CRC64;

Query Match 19.3%; Score 90; DB 3; Length 750;
Best Local Similarity 28.2%; Pred. No. 6.3;
Matches 24; Conservative 18; Mismatches 23; Indels 20; Gaps 3

QY 8 QEAG--NFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRQEAANKQKQQLD 65
DB 550 KQAGENHYNLSGSDYETQIKSLESL-----TNSQAEVSPQEKINELNSQID 597
QY 66 EISTNRQAGVOYSGRADEEQOALS 90
DB 598 ELKLNKNEANKY-----QELALS 616

RESULT 7
O53692
ID O53692 PRELIMINARY; PRT; 97 AA.
AC O53692;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 06, Last annotation update)
DE Hypothetical 9.8 kDa protein (PE family protein).
GN RV0287 OR MT035.15 OR MT0300.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL021930; CAA17362.1; -
DR ENBL; AS008937; AAK44524.1; -
DR TIGR; MT0300; -
DR Tuberculist; RV0287; -
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 97 AA; 9778 MW; 927527DA610A1637 CRC64;

Query Match 18.8%; Score 87.5; DB 16; Length 97;
Best Local Similarity 34.9%; Pred. No. 1.1;
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1

RN
RC SEQUENCE FROM N.A.
RP STRAIN=972h-;
RA Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL009227; CAA15821.1; -
DR GeneDB Spombe; SPAC27D7.02C; -
DR InterPro; IPR000237; GRIP_domain.
DR Pfam; PF01465; GRIP; 1.
SQ SEQUENCE 750 AA; 87265 MW; A469AD95C5787042 CRC64;

Query Match 19.3%; Score 90; DB 3; Length 750;
Best Local Similarity 28.2%; Pred. No. 6.3;
Matches 24; Conservative 18; Mismatches 23; Indels 20; Gaps 3

QY 8 QEAG--NFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRQEAANKQKQQLD 65
DB 550 KQAGENHYNLSGSDYETQIKSLESL-----TNSQAEVSPQEKINELNSQID 597
QY 66 EISTNRQAGVOYSGRADEEQOALS 90
DB 598 ELKLNKNEANKY-----QELALS 616

RESULT 7
O53692
ID O53692 PRELIMINARY; PRT; 97 AA.
AC O53692;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 06, Last annotation update)
DE Hypothetical 9.8 kDa protein (PE family protein).
GN RV0287 OR MT035.15 OR MT0300.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL021930; CAA17362.1; -
DR ENBL; AS008937; AAK44524.1; -
DR TIGR; MT0300; -
DR Tuberculist; RV0287; -
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 97 AA; 9778 MW; 927527DA610A1637 CRC64;

Query Match 18.8%; Score 87.5; DB 16; Length 97;
Best Local Similarity 34.9%; Pred. No. 1.1;
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1

RN
RC SEQUENCE FROM N.A.
RP STRAIN=972h-;
RA Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL009227; CAA15821.1; -
DR GeneDB Spombe; SPAC27D7.02C; -
DR InterPro; IPR000237; GRIP_domain.
DR Pfam; PF01465; GRIP; 1.
SQ SEQUENCE 750 AA; 87265 MW; A469AD95C5787042 CRC64;

Query Match 19.3%; Score 90; DB 3; Length 750;
Best Local Similarity 28.2%; Pred. No. 6.3;
Matches 24; Conservative 18; Mismatches 23; Indels 20; Gaps 3

QY 8 QEAG--NFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRQEAANKQKQQLD 65
DB 550 KQAGENHYNLSGSD
```

```

QY 13 PERISGLKTIQIDQVESTAGSLQGWGAGTAQAQAAVVFQEAANKQKQELDEISTNIR 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 FAKAGLMRHTIGQAEQAAQSAQAFHQGESAAAFQAAHARFVAAAKVNTLLDVAQANLG 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 QAGVQYSEADBEQQQALSSQMGF 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 EAAGTYVAAD---AAAASYTGF 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
O53267 PRELIMINARY; PRT; 97 AA.
ID O53267;
AC O53267;
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE PE-family protein.
GN RV3020C OR MTV012.34C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; AL021287; CAAL16105.1;
DR TubercuList; RV3020C;
KW Complete proteome.
SQ SEQUENCE 97 AA; 9842 MW; BA95CB3180EC17F2 CRC64;

Query Match 18.1%; Score 84.5; DB 16; Length 97;
Best Local Similarity 34.9%; Pred. No. 2.1;
Matches 29; Conservative 7; Mismatches 44; Indels 3; Gaps 1;

QY 13 PERISGLKTIQIDQVESTAGSLQGWGAGTAQAQAAVVFQEAANKQKQELDEISTNIR 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 FAKAGLMRHTIGQAEQAAQSAQAFHQGESAAAFQAAHARFVAAAKVNTLLDVAQANLG 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 QAGVQYSEADBEQQQALSSQMGF 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 EAAGTYVAAD---AAAASYTGF 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
O8VJ87 PRELIMINARY; PRT; 102 AA.
ID O8VJ87;
AC O8VJ87;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE PE family protein.
GN MT3105.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

```

```

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007129; AAK47434.1;
DR TIGR; MT3105;
SQ SEQUENCE 102 AA; 10372 MW; 03CB984D6633D2E4 CRC64;

Query Match 18.1%; Score 84.5; DB 16; Length 102;
Best Local Similarity 34.9%; Pred. No. 2.2;
Matches 29; Conservative 7; Mismatches 44; Indels 3; Gaps 1;

QY 13 PERISGLKTIQIDQVESTAGSLQGWGAGTAQAQAAVVFQEAANKQKQELDEISTNIR 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 23 FAKAGLMRHTIGQAEQAAQSAQAFHQGESAAAFQAAHARFVAAAKVNTLLDVAQANLG 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 QAGVQYSEADBEQQQALSSQMGF 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 83 EAAGTYVAAD---AAAASYTGF 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
O921B9 PRELIMINARY; PRT; 1200 AA.
ID O921B9;
AC O921B9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nuclear pore complex-associated protein Tpr (Fragment).
GN TPR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=129/ola;
RA Sandblad L., Hunziker A., Cordes V.C.;
RT "Evolutionarily conserved mouse tpr is a single-copy gene located on
RT chromosome 1.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ298076; CAC40701.1;
DR InterPro; IPR005613; AIP3.
DR Pfam; PF03915; AIP3; 1.
FT NON_TER 1200 1200.
SQ SEQUENCE 1200 AA; 138973 MW; 616663FBA6B267D7 CRC64;

Query Match 18.0%; Score 84; DB 11; Length 1200;
Best Local Similarity 27.1%; Pred. No. 37;
Matches 29; Conservative 16; Mismatches 38; Indels 24; Gaps 4;

QY 2 DAATLAQEGAGNEF-----RISG-----DLKTDQDVESFAGSLQGWGAA 42
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 900 DIATLKQHLNNEAQLASQSTQRTKGQFGDRDDVDDLSQLRAEEQVNDLKERLKTST 959
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 43 GTAQAQ--AAVVFQEAANKQKQELDEISTNIRQAGVQYSEADBEQQQ 87
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 960 SNVEQYRAMVTSLEDSLNKQVTEVHKNIET---VRLKESAEFTQ 1003
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
O77248 PRELIMINARY; PRT; 188 AA.
ID O77248;
AC O77248;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Apolipoprotein-III.
GN APOLP-III.
OS Spodoptera litura.

```

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
 OC Noctuidae; Amphipyrinae; Spodoptera.
 OX NCBI_TaxID=69820;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=99097579; PubMed=9880904;
 RX Kim E., Kim S.H., Choi C.S., Park Y.I., Kim H.R.;
 RT "Cloning and expression of apolipophorin-III from the common cutworm,
 RT Spodoptera litura";
 RL Arch. Insect Biochem. Physiol. 39:166-173(1998).
 DR EMBL; AF04582; AAC63377.1; -
 SQ SEQUENCE 188 AA; 20649 MW; 7A9AE36DF8E432C CRC64;
 Query Match 17.9%; Score 83.5; DB 5; Length 188;
 Best Local Similarity 26.7%; Pred. No. 5.3;
 Matches 27; Conservative 20; Mismatches 45; Indels 9; Gaps 2;
 QY 1 TDA-----ATLAQAGNFERISGDLKTDQIDVESTAGSLQGWGGAAGTAQAQAAVVFQEE 55
 DB 88 TDANAKAKTALFQARQNLEKTAEDLRKSHPDVERQAGELRTKLQAAVQNTAQEVQKLAK 147
 QY 56 RANKOKQELDEISTNRQAGVQYSDREQQ-----QALSSQ 92
 DB 148 VASVETNEKLAFLKAEYENFSKHVEVQKKVHEAASKQ 188
 RESULT 12
 QBED18
 ID Q8ED18 PRELIMINARY; PRT; 914 AA.
 AC Q8ED18;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE Prophase lambdaSO, tail length tape measure protein.
 GN H O S02953.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis";
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AB015734; AAN55966.1; -
 DR TIGR; S02953; -
 KW Complete proteome.
 SQ SEQUENCE 914 AA; 98450 MW; 30189C1C218EDCD9 CRC64;
 Query Match 17.4%; Score 81; DB 16; Length 914;
 Best Local Similarity 31.2%; Pred. No. 51;
 Matches 20; Conservative 14; Mismatches 24; Indels 6; Gaps 1;
 QY 24 IDQVESTAGSLQGWGGAAGTAQAQAAVVFQEAANKQKQELDEISTNRQAGVQYSDRE 83
 DB 11 LNNVAETGSFVEGQ-----TKAEALAKTEAAAKQKAEINRLGLDPLVAYEYKLDK 64
 QY 84 EQQQ 87
 DB 65 MEQQ 68

RESULT 13
 Q9FSN1
 ID Q9FSN1 PRELIMINARY; PRT; 707 AA.
 AC Q9FSN1;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Methyl-accepting chemotaxis protein McpU.
 GN MGPU
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Maschler P.F.J., Scharf B., Schmitt R.;
 RT "Methyl-accepting chemotaxis protein McpU of Sinorhizobium meliloti";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF12876; AAG34154.1; -
 DR HSP; P02942; IQU7.
 DR InterPro; IPR004010; Cache.
 DR InterPro; IPR004089; Chmtaxis_transd.
 DR InterPro; IPR003660; HAMP.
 DR InterPro; IPR003122; TarH.
 DR Pfam; PF02743; Cache; 1.
 DR Pfam; PF00672; HAMP; 1.
 DR Pfam; PF00015; MCPsignal; 1.
 DR Pfam; PF02203; TarH; 1.
 DR SMART; SMO0304; HAMP; 1.
 DR SMART; SMO0283; MA; 1.
 DR PROSITE; PS00111; CHEMOTAXIS_TRANSDUC_2; 1.
 SQ SEQUENCE 707 AA; 74391 MW; A82D7762FD81FFCC CRC64;
 Query Match 17.3%; Score 80.5; DB 2; Length 707;
 Best Local Similarity 31.1%; Pred. No. 42;
 Matches 33; Conservative 11; Mismatches 39; Indels 23; Gaps 4;
 QY 4 ATLAQAGNFERISGDLKTDQIDVESTAG-----SLOGWGAAGTAQAQAAVVFQEE 55
 DB 570 AVVAQEVRELAQSAANAKEIKALITTSQHGSGVTLVGD---TCRALQAIWVEYQEI 625
 QY 56 -----AANKQKQELDEIST--NIRQAGVQYSDREQQQALS 90
 DB 626 NKHVSALVITATREQSTGLQEIINTAVNTMDQGTQONAAWVEQQTAAS 671
 RESULT 14
 Q26589
 ID Q26589 PRELIMINARY; PRT; 528 AA.
 AC Q26589;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Myosin II heavy chain (Fragment).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Puerto Rican;
 RC Amory L.M.;
 RA Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Puerto Rican;
 RC MEDLINE=93056536; PubMed=1431131;
 RA Soisson L.M., Masterson C.P., Tom T.D., McNally M.T., Lowell G.H.,
 RA Strand M.;
 RT "Induction of protective immunity in mice using a 62-kDa recombinant
 RT fragment of a Schistosoma mansoni surface antigen";
 RL J. Immunol. 149:3612-3620(1992).
 DR EMBL; X65591; CAA46548.1; -

```
DR HSP; P24733; LWDC.
DR InterPro; IPR001637; GlnA_adenyln.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; Myosin_head.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR SMART; SM00015; IQ; 1.
DR PROSITE; PS00182; GLNA ADENYLATION; 1.
DR PROSITE; PS50096; IQ; 1.
FT NON_TER 1
SQ SEQUENCE 528 AA; 61622 MW; AF075D13EB249B4C CRC64;

Query Match 17.2%; Score 80; DB 5; Length 528;
Best Local Similarity 22.8%; Pred. No. 34;
Matches 21; Conservative 21; Mismatches 44; Indels 6; Gaps 2;

QY 2 DAATLAGEAGNFERISGDLKTQIDQVESTAGSLQGWRGAGTAQAQAAVVRFOEAANKQK 61
DB 341 DLKATQETVDDLERVKRDLEQLRRKEAEIGGLSGKFDEQGLVAQ-----LQRKIKELQ 1098
QY 62 QELDEISTNIRQAGVQYSRADEEQOQALSSQM 93
DB 396 TRIQEEEDLEAERAARSKAEKSRQO-LESEL 426

RESULT 15
Q02456 PRELIMINARY; PRT; 1940 AA.
AC Q02456;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Myosin heavy chain.
GN MYH.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Brazilian LE;
RX MEDLINE=93211444; PubMed=8459827;
RA Weston D.S., Schmitz J., Kemp M., Kunz W.;
RT "Cloning and sequence characterization of a complete myosin heavy
RL chain cDNA from Schistosoma mansoni.";
RL Mol. Biochem. Parasitol. 58:161-164 (1993).
DR EMBL; L01634; AAA2905.1; -.
DR HSP; P08799; LMND.
DR InterPro; IPR001637; GlnA_adenyln.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001603; Myosin_head.
DR InterPro; IPR004003; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00194; TROPOMYOSIN.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00182; GLNA ADENYLATION; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 1940 AA; 222379 MW; 10FC4EAE208CA365 CRC64;

Query Match 17.2%; Score 80; DB 5; Length 1940;
Best Local Similarity 22.8%; Pred. No. 1.4e+02;
```

```
Matches 21; Conservative 21; Mismatches 44; Indels 6; Gaps 2;
QY 2 DAATLAGEAGNFERISGDLKTQIDQVESTAGSLQGWRGAGTAQAQAAVVRFOEAANKQK 61
DB 1044 DLKATQETVDDLERVKRDLEQLRRKEAEIGGLSGKFDEQGLVAQ-----LQRKIKELQ 1098
QY 62 QELDEISTNIRQAGVQYSRADEEQOQALSSQM 93
DB 1099 TRIQEEEDLEAERAARSKAEKSRQO-LESEL 1129

Search completed: November 21, 2003, 16:08:11
Job time : 5.6746 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 23.6728 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-26

Perfect score: 1949

Sequence: 1 MYDFGALPPEINSARYAGP.....SGVLRVPRPYVMPHSPAG 391

Scoring table: BLOSUM62

Gapop:10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```
1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1949	100.0	391	AAW32449	Mycobacterium tube
2	1949	100.0	391	AAW32381	Mycobacterium tube
3	1949	100.0	391	AAW81702	M. tuberculosis im
4	1949	100.0	391	AAW64335	Mycobacterium tube
5	1949	100.0	391	AA33132	M. tuberculosis an
6	1949	100.0	391	AA33989	M. tuberculosis re
7	1949	100.0	391	AA04778	Mycobacterium spec
8	1949	100.0	391	AAU01888	M. tuberculosis an
9	1949	100.0	391	AAE29707	Mycobacterium sp.

10	1949	100.0	391	23	AAE17571	Mycobacterium spec
11	1949	100.0	596	20	AAV32070	Mycobacterium tube
12	1949	100.0	596	23	AAE29710	Mycobacterium sp.
13	1949	100.0	595	23	AAE17574	Mycobacterium spec
14	1949	100.0	599	23	AAU74599	Antigenic fusion p
15	1949	100.0	600	20	AAV32068	Mycobacterium tube
16	1949	100.0	600	23	AAU74597	Antigenic fusion p
17	1949	100.0	729	22	AAO22142	Ral12-H9-32A fusion
18	1949	100.0	729	23	AAE29709	Mycobacterium spec
19	1949	100.0	729	23	AAE17573	Mycobacterium sp.
20	1949	100.0	744	22	AAU01902	M. tuberculosis an
21	1949	100.0	815	22	AAU01904	M. tuberculosis an
22	1949	100.0	930	23	AAE29731	Mycobacterium sp.
23	1945	99.8	788	22	AAU01903	M. tuberculosis an
24	1944	99.7	394	20	AAV04779	Mycobacterium spec
25	1944	99.7	729	23	AAE29708	Mycobacterium spec
26	1944	99.7	729	23	AAE17572	Mycobacterium spec
27	1931	99.1	729	20	AAV32059	Mycobacterium tube
28	1902.5	97.6	726	23	AAU74588	Antigenic fusion p
29	1652.5	84.8	396	19	AAW81704	M. tuberculosis im
30	1652.5	84.8	396	19	AAW64337	Mycobacterium tube
31	1652.5	84.8	396	20	AAV39134	M. tuberculosis an
32	1652.5	84.8	396	20	AAV38991	M. tuberculosis re
33	1486.5	76.3	359	19	AAW81703	M. tuberculosis im
34	1486.5	76.3	359	19	AAW64336	Mycobacterium tube
35	1486.5	76.3	359	20	AAV39133	M. tuberculosis an
36	1486.5	76.3	359	20	AAV38990	M. tuberculosis re
37	1474.5	75.7	341	18	AAW32451	Mycobacterium tube
38	1474.5	75.7	341	18	AAW32383	Mycobacterium tube
39	1187.5	60.9	358	23	AAU74591	Antigenic fusion p
40	1187	60.9	263	18	AAW32447	Mycobacterium tube
41	1187	60.9	263	18	AAW32379	Mycobacterium tube
42	1187	60.9	263	19	AAW81680	M. tuberculosis im
43	1187	60.9	263	19	AAW64317	Mycobacterium tube
44	1187	60.9	263	20	AAV32062	Mycobacterium tube
45	1187	60.9	263	20	AAV39121	M. tuberculosis an

ALIGNMENTS

RESULT 1
AAW32449

ID AAW32449 standard; Protein; 391 AA.

AC AAW32449;

XX 09-JAN-1998 (first entry)

DT Mycobacterium tuberculosis antigen TbH-9FL.

DE Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

OS W09709428-A2.

FN 13-MAR-1997.

PD 30-AUG-1996; 96WO-US14674.

XX 12-JUL-1996; 96US-0680574.

PR 01-SEP-1995; 95US-0523436.

PR 22-SEP-1995; 95US-0533634.

PR 22-MAR-1996; 96US-0620874.

PR 05-JUN-1996; 96US-0659683.

XX (CORI-) CORIXA CORP.

PA Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

XX Twardzik DR, Vedrick TH;

PI

DR WPI; 1997-192903/17.
 DR N-PSDB; AAT91521.
 XX
 PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also
 PT for diagnosis
 XX
 PS Example 3; Page 138-139; 168pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, Tbh-9FL The immunogenic protein, and fusion proteins
 CC containing one or more of the proteins or one of the proteins plus
 CC EGAT-6, are useful in vaccines, preferably when formulated with a
 CC non-specific adjuvant, to induce an immune response against
 CC M.tuberculosis (for treatment or prevention).
 XX
 SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 18; Length 391;
 Best Local Similarity 100.0%; Pred. No. 4.8e-143;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 Db 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 Qy 61 SSAGLMVAASPPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 Db 61 SSAGLMVAASPPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 Qy 121 LIATNLLGQNTPAIVNAAEYEGEMWADAAAMFGYAAATATATATLLPFEAPMTSAGG 180
 Db 121 LIATNLLGQNTPAIVNAAEYEGEMWADAAAMFGYAAATATATATLLPFEAPMTSAGG 180
 Qy 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTOGTTPSSKLGWLKTVSPHRSPIIN 240
 Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTOGTTPSSKLGWLKTVSPHRSPIIN 240
 Qy 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSSLGSSG 300
 Db 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSSLGSSG 300
 Qy 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGFQMLGGLPV 360
 Db 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGFQMLGGLPV 360
 Qy 361 GOMGARAGGGLSGVLRVPPPPYVMPHSPAAG 391
 Db 361 GOMGARAGGGLSGVLRVPPPPYVMPHSPAAG 391
 RESULT 2
 AAW32381
 ID AAW32381 standard; Protein; 391 AA.
 XX
 AC AAW32381;
 XX
 DT 13-JAN-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Tbh-9FL.
 XX
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 XX skin testing; M.tuberculosis.
 XX
 CS Mycobacterium tuberculosis.
 XX
 FN WO9709429-A2.
 XX
 PD 13-MAR-1997.
 XX

PF 30-AUG-1996; 96WO-US14675.
 XX
 PR 12-JUL-1996; 96US-0680573.
 PR 01-SEP-1995; 95US-0523435.
 PR 22-SEP-1995; 95US-0532136.
 PR 22-MAR-1996; 96US-0620280.
 PR 05-JUN-1996; 96US-0658800.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
 PI Twardzik DR, Vedvick TH;
 XX
 DR WPI; 1997-192904/17.
 DR N-PSDB; AAT91455.
 XX
 PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
 PT - useful for diagnosis of M. tuberculosis infection
 XX
 XX Example 3; Page 150-152; 190pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, Tbh-9FL The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX
 SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 18; Length 391;
 Best Local Similarity 100.0%; Pred. No. 4.8e-143;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 Db 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 Qy 61 SSAGLMVAASPPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 Db 61 SSAGLMVAASPPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 Qy 121 LIATNLLGQNTPAIVNAAEYEGEMWADAAAMFGYAAATATATATLLPFEAPMTSAGG 180
 Db 121 LIATNLLGQNTPAIVNAAEYEGEMWADAAAMFGYAAATATATATLLPFEAPMTSAGG 180
 Qy 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTOGTTPSSKLGWLKTVSPHRSPIIN 240
 Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTOGTTPSSKLGWLKTVSPHRSPIIN 240
 Qy 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSSLGSSG 300
 Db 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSSLGSSG 300
 Qy 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGFQMLGGLPV 360
 Db 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGFQMLGGLPV 360
 Qy 361 GOMGARAGGGLSGVLRVPPPPYVMPHSPAAG 391
 Db 361 GOMGARAGGGLSGVLRVPPPPYVMPHSPAAG 391
 RESULT 3
 AAW81702
 ID AAW81702 standard; Protein; 391 AA.
 XX
 AC AAW81702;
 XX

```

DT 27-JAN-1999 (first entry)
XX M. tuberculosis immunogenic polypeptide TbH-9FL.
DE Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX Mycobacterium tuberculosis.
OS
XX WO9816646-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18293.
XX
XX 13-MAR-1997; 97US-0818112.
XX 11-OCT-1996; 96US-0730510.
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1998-261042/23.
XX N-PSDB; AAV64503.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
XX
XX Example 3B; Page 128-129; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
CC for inducing protective immunity against tuberculosis (TB). This
CC sequence can be formulated into vaccines and/or pharmaceutical
CC compositions for immunising against M. tuberculosis infection or may
CC be used for the diagnosis of tuberculosis.
XX
XX Sequence 391 AA;
SQ
    Query Match          100.0%; Score 1949; DB 19; Length 391;
    Best Local Similarity 100.0%; Pred. No. 4.8e-143;
    Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAAASPYVAMSVTAQOAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMSVTAQOAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPEMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPEMTSAGG 180
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQOQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEASDTAAANQLMNNVPQALQOQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAQAQVOTAAQNGVRAMWSLGSLSGSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAQAQVOTAAQNGVRAMWSLGSLSGSG 300
QY 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAAEERGPGQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAAEERGPGQMLGGLPV 360
QY 361 GQMGARAGGLSGVLVRRPVPYVMPHSPAAG 391
DB 361 GQMGARAGGLSGVLVRRPVPYVMPHSPAAG 391

```

```

RESULT 4
AAW64335
ID AAW64335 standard; Protein; 391 AA.
XX
XX AAW64335;
AC
XX
XX 09-NOV-1998 (first entry)
DT
XX Mycobacterium tuberculosis antigen TbH-9FL.
DE
XX Tuberculosis; infection; diagnosis; antigen; TbH-9FL.
KW
XX Mycobacterium tuberculosis strain H37Rv.
OS
XX WO9816645-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18214.
XX
XX 13-MAR-1997; 97US-0818111.
XX 11-OCT-1996; 96US-0729622.
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1998-251292/22.
XX N-PSDB; AAV44395.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
XX Example 3; Page 133-135; 250pp; English.
XX
XX This polypeptide comprises Mycobacterium tuberculosis antigen
CC TbH-9FL. It is encoded by genomic DNA (see AAV44395) isolated from a
CC M. tuberculosis strain H37Rv genomic library using a probe from
CC clone TbH-9 (see AAV44371). The invention relates to compositions
CC and methods for diagnosing tuberculosis. It provides polypeptides
CC (see AAW64291-W64379) comprising an antigenic portion of a soluble
CC M. tuberculosis antigen, or an immunogenic portion of an M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers, for the diagnosis of tuberculosis.
XX
XX Sequence 391 AA;
SQ
    Query Match          100.0%; Score 1949; DB 19; Length 391;
    Best Local Similarity 100.0%; Pred. No. 4.8e-143;
    Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAAASPYVAMSVTAQOAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMSVTAQOAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPEMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPEMTSAGG 180
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQOQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEASDTAAANQLMNNVPQALQOQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240

```

QY 241 MVSNNHMSMTNSGVSMTNTLSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 300
 DB 241 MVSNNHMSMTNSGVSMTNTLSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 300
 QY 301 LGGVAAANLGRAASVGSLSVFPQAWAANAQAVTPAARALPLTSLTSAERGPQGMGLGLPV 360
 DB 301 LGGVAAANLGRAASVGSLSVFPQAWAANAQAVTPAARALPLTSLTSAERGPQGMGLGLPV 360
 QY 361 GQMGARAGGSLGSLRVPPRPYVMPHSPAAG 391
 DB 361 GQMGARAGGSLGSLRVPPRPYVMPHSPAAG 391
 RESULT 5
 AAY39132
 ID AAY39132 standard; Protein; 391 AA.
 AC AAY39132;
 XX
 DT 05-NOV-1999 (first entry)
 DE M. tuberculosis antigen TbH-9FL amino acid sequence.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX WO9942076-A2.
 XX
 XX 26-AUG-1999.
 XX
 XX 17-FEB-1999; 99WO-US03268.
 XX
 XX 05-MAY-1998; 98US-0072967.
 XX 18-FEB-1998; 98US-0025197.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 XX WPI; 1999-527409/44.
 DR N-PSDB; AAZ19305.
 XX
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 XX
 XX Example 3; Page 123-124; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.
 XX
 XX Sequence 391 AA;
 SQ
 Query Match 100.0%; Score 1949; DB 20; Length 391;
 Best Local Similarity 100.0%; Pred. NO. 4.8e-143;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDGALPPEINARMYAGPQSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
 DB 1 MVDGALPPEINARMYAGPQSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60

QY 51 SSAGLWVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAVETAYGLTVPPVIAENRAELMI 120
 DB 51 SSAGLWVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAVETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLGONTPTAIJAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 DB 121 LIATNLGONTPTAIJAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 QY 181 LLEQAAVEEASDTAAANQNLMMNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHRSISN 240
 DB 181 LLEQAAVEEASDTAAANQNLMMNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHRSISN 240
 QY 241 MVSNNHMSMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 DB 241 MVSNNHMSMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 QY 301 LGGVAAANLGRAASVGSLSVFPQAWAANAQAVTPAARALPLTSLTSAERGPQGMGLGLPV 360
 DB 301 LGGVAAANLGRAASVGSLSVFPQAWAANAQAVTPAARALPLTSLTSAERGPQGMGLGLPV 360
 QY 361 GQMGARAGGSLGSLRVPPRPYVMPHSPAAG 391
 DB 361 GQMGARAGGSLGSLRVPPRPYVMPHSPAAG 391
 RESULT 6
 AAY38989
 ID AAY38989 standard; Protein; 391 AA.
 AC AAY38989;
 XX
 DT 05-NOV-1999 (first entry)
 DE M. tuberculosis recombinant antigen protein TbH-9FL.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX WO9942118-A2.
 XX
 XX 26-AUG-1999.
 XX
 XX 17-FEB-1999; 99WO-US03285.
 XX
 XX 05-MAY-1998; 98US-0072596.
 XX 18-FEB-1998; 98US-0024753.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 XX WPI; 1999-527416/44.
 DR N-PSDB; AAZ19093.
 XX
 XX New polypeptide comprising antigenic portions of M. tuberculosis
 PT Example 3; Page 168-169; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX
 XX Sequence 391 AA;
 SQ
 Query Match 100.0%; Score 1949; DB 20; Length 391;

Example 2; Page 151-152; 168pp; English.

PS The sequence represents Mycobacterium tuberculosis TbH9 (also known
XX as Mtb39A), an M. tuberculosis antigen. Compositions comprising at least
CC 2 heterologous antigens, as a fusion protein, and vectors expressing the
CC fusion proteins are used as vaccines to prophylactically immunise
CC mammals (especially humans) against infection by Mycobacterium.
CC Compositions contain at least 2 heterologous antigens that increase the
CC serological sensitivity of individuals infected with tuberculosis, a
CC disease frequently affecting patients with acquired immunodeficiency
XX disease, AIDS.

XX Sequence 391 AA;
SQ Query Match 100.0%; Score 1949; DB 22; Length 391;
Best Local Similarity 100.0%; Pred. No. 4.8e-143;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVDGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSWIG 60
Db 1 MVDGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSWIG 60
Qy 61 SSAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Qy 121 LIATNLGQNTPAIVNEAEYGEVMAQDAAMFGYAAATATATATATLLPFEAPMTSAGG 180
Db 121 LIATNLGQNTPAIVNEAEYGEVMAQDAAMFGYAAATATATATLLPFEAPMTSAGG 180
Qy 181 LLEQAAVEEASDTAAANQLMNNVPQALQLOAQTOGTPSSKLGGLWKTVPSPHRSPI 240
Db 181 LLEQAAVEEASDTAAANQLMNNVPQALQLOAQTOGTPSSKLGGLWKTVPSPHRSPI 240
Qy 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 300
Db 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 300
Qy 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
Qy 361 GQMGARAGGGLSGVLVRPPPPVYVMPHSPAAG 391
Db 361 GQMGARAGGGLSGVLVRPPPPVYVMPHSPAAG 391

RESULT 9
AAE29707
ID AAE29707 standard; Protein: 391 AA.
XX AC AAE29707;
XX DT 27-JAN-2003. (first entry)
XX DE Mycobacterium sp. TbH9FL antigenic protein.
XX KW Vaccine; immunity; diagnostic agent; gene therapy; TbH9FL antigen.
XX OS Mycobacterium sp.
XX WO200272792-A2.
XX PD 19-SEP-2002.
XX PF 13-MAR-2002; 2002WO-US08223.
XX PR 13-MAR-2001; 2001US-275837P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Brannon M, Guderian J;
XX

DR WPI; 2002-753844/82.
DR N-PSDB; AAD47082.
XX New recombinant nucleic acid molecule comprising a Leishmania TSA,
PT LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective
PT immunity against pathogenic microorganisms e.g. Leishmania and
XX Mycobacterium tuberculosis
PS Disclosure; Page 86-87; 155pp; English.
XX The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides,
CC as in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC Mycobacterium sp. TbH9FL antigenic protein.
XX Sequence 391 AA;
SQ Query Match 100.0%; Score 1949; DB 23; Length 391;
Best Local Similarity 100.0%; Pred. No. 4.8e-143;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVDGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSWIG 60
Db 1 MVDGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSWIG 60
Qy 61 SSAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Qy 121 LIATNLGQNTPAIVNEAEYGEVMAQDAAMFGYAAATATATATLLPFEAPMTSAGG 180
Db 121 LIATNLGQNTPAIVNEAEYGEVMAQDAAMFGYAAATATATATLLPFEAPMTSAGG 180
Qy 181 LLEQAAVEEASDTAAANQLMNNVPQALQLOAQTOGTPSSKLGGLWKTVPSPHRSPI 240
Db 181 LLEQAAVEEASDTAAANQLMNNVPQALQLOAQTOGTPSSKLGGLWKTVPSPHRSPI 240
Qy 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 300
Db 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 300
Qy 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
Qy 361 GQMGARAGGGLSGVLVRPPPPVYVMPHSPAAG 391
Db 361 GQMGARAGGGLSGVLVRPPPPVYVMPHSPAAG 391

RESULT 10
AAE17571
ID AAE17571 standard; Protein: 391 AA.
XX AC AAE17571;
XX DT 22-APR-2002 (first entry)
XX DE Mycobacterium species MTB39 (TbH9) protein #2.
XX KW Fusion protein; antigen; serological sensitivity; immune response;
XX tuberculosis; infection; vaccine; MTB39; TbH9 protein.
XX OS Mycobacterium sp.

XX W0200198460-A2.
XX 27-DEC-2001.
XX 20-JUN-2001; 2001WO-US19959.
XX 20-JUN-2000; 2000US-0597796.
XX 01-FEB-2001; 2001US-2657372.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Reed S, Alderson M;
XX WPI; 2002-147798/19.
XX N-PSDB; RAD28341.
XX Composition comprising MTB39 antigen and MTB32A antigen from
XX Mycobacterium species, useful for eliciting immune response in a
XX subject -
XX Claim 83; Page 102-103; 136pp; English.
XX The present invention relates to fusion proteins containing at least
XX two Mycobacterium species antigens, nucleotides encoding them and
XX compositions comprising such fusion proteins. The present invention
XX particularly relates to nucleic acids encoding fusion proteins that
XX include two or more individual M. tuberculosis antigens which increase
XX the serological sensitivity of sera from individuals infected with
XX tuberculosis and methods for their use in diagnosis, prevention and
XX treatment of tuberculosis infection. Sequences of the invention are
XX useful for eliciting an immune response in a mammal, e.g., human,
XX immunised with BCG. They are useful in the diagnosis, treatment and
XX prevention of Mycobacterium infection. The fusion proteins and the
XX polynucleotides are useful as diagnostic tools in patients infected
XX with Mycobacterium, in vitro and in vivo assays for detecting humoral
XX antibodies or cell-mediated immunity against M. tuberculosis, for the
XX diagnosis of an infection or monitoring of disease progression, as
XX immunogens to generate or elicit a protective immune response in a
XX patient and for raising anti-M. tuberculosis antibodies in a non-human
XX animal. Sequences of the invention are also used as vaccines. MTB32A
XX fusion proteins of the invention are useful as in vivo diagnostic agents
XX for intradermal skin test. The present sequence is Mycobacterium species
XX MTB39 (TBH9) protein.
XX Sequence 391 AA;
Query Match 100.0%; Score 1949; DB 23; Length 391;
Best Local Similarity 100.0%; Pred. No. 4.8e-143;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPPEINSGARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSWIG 60
DB 1 MVDGALPPPEINSGARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSWIG 60
QY 61 SSAGLWVAASPYVWMSVTAGOAEHTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
DB 61 SSAGLWVAASPYVWMSVTAGOAEHTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
QY 121 LIATNLGQNTPALVAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLGQNTPALVAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEASDTAAANQNLNNVPQALQOLAQTGQTTPSSKGLGKWTVPSPHSPISN 240
DB 181 LLEQAAVEASDTAAANQNLNNVPQALQOLAQTGQTTPSSKGLGKWTVPSPHSPISN 240
QY 241 MVSNNHMSMTNSGVSMNTLTSSMLKGFAPAAAQAVQTAAGNGVRAMSSLGSSSG 300
DB 241 MVSNNHMSMTNSGVSMNTLTSSMLKGFAPAAAQAVQTAAGNGVRAMSSLGSSSG 300
QY 301 LGGVAANLGRAASVGSLSVPOAWAANQAVTTPAARALPLTSLTSAERGPQMLGGLPV 360

DB 301 LGGVAANLGRAASVGSLSVPOAWAANQAVTTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPRPYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLRVPRPYVMPHSPAAG 391
RESULT 11
AAY32070
ID AAY32070 standard; Protein; 596 AA.
XX AAY32070;
XX 17-JAN-2000 (first entry)
XX Mycobacterium tuberculosis antigen fusion protein Mtb59f.
XX Tuberculosis; antigen; fusion protein; Mtb59f; TBH9; Ra35;
XX diagnosis; therapy; vaccine; immunogen.
XX Mycobacterium tuberculosis.
XX Key Location/Qualifiers
FT Peptide 1..8
FT Protein /note= "Met/His tag"
FT Protein 9..140
FT Protein /note= "Ra12"
FT Protein 143..596
FT Protein /note= "TBH9"
XX W09951748-A2.
XX 14-OCT-1999.
XX 07-APR-1999; 99WO-US07717.
XX 07-APR-1999; 98US-0056556.
XX 30-DEC-1999; 98US-0223040.
XX (CORI-) CORIXA CORP.
XX Skeiky YAW, Alderson M, Campos-Neto A;
XX WPI; 1999-601610/51.
XX N-PSDB; AA220205.
XX New fusion proteins useful for diagnosis, prevention and treatment of
XX tuberculosis -
XX Claim 1; Fig 12A-B; 83pp; English.
XX This sequence represents a recombinant Mycobacterium tuberculosis
XX bi-antigen fusion protein, termed Mtb59f, composed of the antigens
XX TBH9 and Ra35. The fusion protein is expressed in host cells
XX using a vector carrying a polynucleotide (see AA220205) comprising
XX the coding sequences for the 2 antigens. The invention provides
XX fusion proteins (see AAY32059-71) containing at least 2 M.
XX tuberculosis antigens. The new fusion proteins and polynucleotides
XX encoding them are useful as vaccines for preventing tuberculosis
XX (claimed), for diagnosis (via in vitro assays or intradermal skin
XX tests for detection of anti-M. tuberculosis antibodies), monitoring
XX of disease progression, and treatment of tuberculosis. They are
XX more effective immunogens than mixtures of the individual protein
XX components.
XX Sequence 596 AA;
Query Match 100.0%; Score 1949; DB 20; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.1e-143;
Matches 591; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPPEINSGARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSWIG 60

Db 9 MVDFCALPPEINARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
 QY 61 SSAGLMVAAAAPYVWMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 69 SSAGLMVAAAAPYVWMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPEMTSAGG 180
 Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 188
 QY 181 LLEQAAAEEASDTAAANQLMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHSPISN 240
 Db 189 LLEQAAAEEASDTAAANQLMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHMTNSGVSWTNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 300
 Db 249 MYSMANNHMTNSGVSWTNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 308
 QY 301 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
 Db 309 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
 Db 369 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 399

RESULT 12
 AAE29710
 ID AAE29710 standard; Protein; 596 AA.
 XX
 AC AAE29710;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium sp. MTB59F fusion protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen;
 KW Ra35; MTB59F; fusion protein.
 XX
 OS Chimeric - Mycobacterium sp.
 OS Chimeric - Mycobacterium tuberculosis.
 XX
 PN WO200272792-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 13-MAR-2002; 2002WO-US08223.
 XX
 PR 13-MAR-2001; 2001US-275837P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Brannon M, Guderian J;
 XX
 DR WPI; 2002-759844/82.
 DR N-PSDB; AAD47086.
 XX
 PT New recombinant nucleic acid molecule comprising a Leishmania TSA.
 PT Leif, M15 or 6H polynucleotide, useful as vaccine to elicit protective
 PT immunity against pathogenic microorganisms e.g. Leishmania and
 PT Mycobacterium tuberculosis
 XX
 PS Disclosure; Page 98-99; 155pp; English.
 XX
 CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic

CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides,
 CC as in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC MTB59F fusion protein. This fusion protein comprises Ra35 protein from
 CC Mycobacterium tuberculosis and TBH9 protein from Mycobacterium sp.
 XX
 SQ Sequence 596 AA;
 Query Match 100.0%; Score 1949; DB 23; Length 596;
 Best Local Similarity 100.0%; Pred. No. 8.1e-143;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDFGALPPEINARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
 Db 9 MVDFGALPPEINARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
 QY 61 SSAGLMVAAAAPYVWMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 69 SSAGLMVAAAAPYVWMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
 Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 188
 QY 181 LLEQAAAEEASDTAAANQLMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHSPISN 240
 Db 189 LLEQAAAEEASDTAAANQLMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHMTNSGVSWTNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 300
 Db 249 MYSMANNHMTNSGVSWTNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 308
 QY 301 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
 Db 309 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
 Db 369 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 399

RESULT 13
 AAE17574
 ID AAE17574 standard; Protein; 596 AA.
 XX
 AC AAE17574;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium species MTB59F fusion protein.
 XX
 KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; TBH9-Ra35 protein.
 XX
 OS Mycobacterium sp.
 XX
 PN WO200198460-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 20-JUN-2001; 2001WO-US19959.
 XX
 PR 20-JUN-2000; 2000US-0597796.
 PR 01-FEB-2001; 2001US-265737P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Alderson M;
 XX
 DR WPI; 2002-147798/19.
 DR N-PSDB; AAD28344.
 XX

PT Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a
XX subject

XX Claim 5; Page 114-115; 136pp; English.

XX The present invention relates to fusion proteins containing at least
CC two Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected
CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC MTB59F (TbH9-Ra35) fusion protein.

XX Sequence 596 AA;

Query Match 100.0%; Score 1949; DB 23; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.1e-143;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGSWIG 60
Db 9 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGSWIG 68
Qy 61 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 69 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
Qy 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 190
Db 129 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 198
Qy 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
Db 189 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248
Qy 241 MVSMAHHMTNSGVSWNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLGSSG 300
Db 249 MVSMAHHMTNSGVSWNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLGSSG 308
Qy 301 LGGVAAANLGRAASVGLSVFQAWAANQAVTTPAARALPLTSLTSAARPGQWGLGLPV 360
Db 309 LGGVAAANLGRAASVGLSVFQAWAANQAVTTPAARALPLTSLTSAARPGQWGLGLPV 368
Qy 361 GQMGARAGGGLSGVLVPPPPYVPHSPAAG 391
Db 369 GQMGARAGGGLSGVLVPPPPYVPHSPAAG 399

RESULT 14

AAU74599
ID AAU74599 standard; Protein; 599 AA.

XX AC AAU74599;

XX DT 08-MAY-2002 (first entry)

XX Antigenic fusion protein Tb59-Ra35 (Mtb59f).

XX

KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
KW tuberculosatic; immunogen; vaccine; Tb59-Ra35; Mtb59f.

OS Chimeric - Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT Misc-difference 597

FT /label= OTHER
FT /note= "OTHER= Xaa. Xaa= In frame stop codon"

XX US2002009459-A1.

XX 24-JAN-2002.

XX 07-APR-1999; 99US-0287849.

XX 13-MAR-1997; 97US-0818112.

XX 18-OCT-1997; 97US-0943578.

XX 18-FEB-1998; 98US-0025197.

XX 07-APR-1998; 98US-0056556.

XX 30-DEC-1998; 98US-0223040.

XX (REED/) REED S G.

XX (SKEL/) SKEIKY Y A.

XX (DILL/) DILLON D C.

XX (ALDE/) ALDERSON M.

XX (CAMF/) CAMPOS-NETO A.

XX Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;

XX WPI; 2002-171134/22.

XX N-PSDB; ABK14139.

XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
PT diagnosing, treating or preventing M. tuberculosis infection,
PT particularly as vaccine for treating or preventing tuberculosis

XX Claim 1; Fig 12; 62pp; English.

XX The invention relates to a purified polypeptide which induces an immune
CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
CC useful for diagnosing, treating or preventing M. tuberculosis infection,
CC particularly tuberculosis infection. In particular, the polypeptides are
CC useful as a vaccine formulation with an adjuvant to afford long-term
CC protection in animals against the development of tuberculosis. The
CC protein coding sequence may be used to encode a protein product for use
CC as an immunogen to induce and/or enhance an immune response to M.
CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
CC of the invention.

XX Sequence 599 AA;

Query Match 100.0%; Score 1949; DB 23; Length 599;
Best Local Similarity 100.0%; Pred. No. 8.1e-143;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGSWIG 60

Db 9 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGSWIG 68

Qy 61 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

Db 69 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128

Qy 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180

Db 129 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188

Qy 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240

Db 189 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248

Qy 241 MVSMAHHMTNSGVSWNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLGSSG 300

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 8.24558 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-26

Perfect score: 1949

Sequence: 1 MVDFGALPEINARMVAGP.....SGVLVPPRPVMPHPSPAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PGTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1949	100.0	391	3	US-08-818-112-107
2	1949	100.0	391	4	US-08-818-111-102
3	1949	100.0	391	4	US-09-056-556-107
4	1949	100.0	391	4	US-09-072-596-102
5	1944	99.7	729	4	US-09-223-040-2
6	1652.5	84.8	396	3	US-08-818-112-111
7	1652.5	84.8	396	4	US-08-818-111-106
8	1652.5	84.8	396	4	US-09-056-556-111
9	1652.5	84.8	396	4	US-09-072-596-106
10	1486.5	76.3	359	3	US-08-818-112-109
11	1486.5	76.3	359	4	US-08-818-111-104
12	1486.5	76.3	359	4	US-09-056-556-109
13	1486.5	76.3	359	4	US-09-072-596-104
14	1187	60.9	263	3	US-08-818-112-91
15	1187	60.9	263	4	US-08-818-111-92
16	1187	60.9	263	4	US-09-056-556-91
17	1187	60.9	263	4	US-09-072-596-92
18	766.5	39.3	400	4	US-09-073-009-126
19	603	30.9	423	4	US-09-073-009-142
20	424.5	21.8	943	4	US-09-477-135A-131
21	424	21.8	141	4	US-09-073-009-15
22	381.5	19.6	204	4	US-08-311-731A-57
23	377.5	19.4	208	4	US-08-311-731A-208
24	314	16.1	368	3	US-08-818-112-114
25	314	16.1	368	4	US-08-818-111-109
26	314	16.1	368	4	US-09-056-556-114
27	314	16.1	368	4	US-09-072-596-109

28 261 13.4 1271 1 US-08-095-734-2 Sequence 2, Appli
29 261 13.4 1271 2 US-08-444-623-2 Sequence 2, Appli
30 261 13.4 1271 3 US-08-471-869-2 Sequence 2, Appli
31 261 13.4 1271 4 US-09-342-563-2 Sequence 2, Appli
32 261 13.4 1271 5 PCT-US94-08267-2 Sequence 2, Appli
33 215 11.0 352 4 US-09-073-009-14 Sequence 14, Appli
34 186.5 9.6 943 4 US-09-056-556-204 Sequence 204, App
35 186.5 9.6 943 4 US-09-072-596-199 Sequence 199, App
36 145 7.4 800 4 US-09-252-991A-20437 Sequence 20437, A
37 143.5 7.4 228 4 US-09-477-135A-128 Sequence 128, App
38 141 7.2 738 3 US-08-864-038A-3 Sequence 3, Appli
39 139.5 7.2 792 2 US-08-678-039A-40 Sequence 40, Appli
40 131.5 6.7 826 4 US-09-894-998A-47 Sequence 47, Appli
41 129.5 6.6 1186 2 US-08-861-464-8 Sequence 8, Appli
42 129.5 6.6 1186 2 US-08-396-001-8 Sequence 8, Appli
43 129.5 6.6 1186 3 US-09-323-433A-9 Sequence 8, Appli
44 128 6.6 731 4 US-09-340-736E-1 Sequence 1, Appli
45 126 6.5 731 2 US-08-911-364-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Szeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818.112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.4.11C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-107

Query Match 100.0%; Score 1949; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 1 MYDFGALPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLWVAASPYVWMSVTAQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWVAASPYVWMSVTAQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGQNTPAIAVNEAEYEMMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
Db 121 LIATNLGQNTPAIAVNEAEYEMMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAOPTQGTTPSSKLGGLWKTVPSPHSPISN 240
Db 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAOPTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
Db 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
QY 361 GQMGARAGGSLGSLVRPPRPPVMPHSPAAG 391
Db 361 GQMGARAGGSLGSLVRPPRPPVMPHSPAAG 391

RESULT 2

US-08-818-111-102
; Sequence 102, Application US/0818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-102

Query Match 100.0%; Score 1949; PB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYDFGALPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 1 MYDFGALPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLWVAASPYVWMSVTAQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWVAASPYVWMSVTAQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGQNTPAIAVNEAEYEMMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
Db 121 LIATNLGQNTPAIAVNEAEYEMMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAOPTQGTTPSSKLGGLWKTVPSPHSPISN 240
Db 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAOPTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
Db 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
QY 361 GQMGARAGGSLGSLVRPPRPPVMPHSPAAG 391
Db 361 GQMGARAGGSLGSLVRPPRPPVMPHSPAAG 391

RESULT 3

US-09-056-556-107
; Sequence 107, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-056-556-107

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINARMYAGPSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 1 MVDGALPPEINARMYAGPSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAAVEASDTAAANQLMNVPOALQOQAOTGTTTSSKLGGLWKTVSPHRSPISN 240
DB 181 LLEQAAAVEASDTAAANQLMNVPOALQOQAOTGTTTSSKLGGLWKTVSPHRSPISN 240

QY 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 300
DB 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLVPRPVYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLVPRPVYVMPHSPAAG 391

RESULT 4

US-09-072-596-102
Sequence 102, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-102

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINARMYAGPSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 1 MVDGALPPEINARMYAGPSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAAVEASDTAAANQLMNVPOALQOQAOTGTTTSSKLGGLWKTVSPHRSPISN 240
DB 181 LLEQAAAVEASDTAAANQLMNVPOALQOQAOTGTTTSSKLGGLWKTVSPHRSPISN 240

QY 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 300
DB 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLVPRPVYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLVPRPVYVMPHSPAAG 391

RESULT 5

US-09-223-040-2
Sequence 2, Application US/09223040
Patent No. 6544522
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009010US
CURRENT APPLICATION NUMBER: US/09/223,040
CURRENT FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 2
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-223-040-2

Query Match 99.7%; Score 1944; DB 4; Length 729;
Best Local Similarity 99.7%; Pred. No. 1.7e-153;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVDGALPPEINARMYAGPSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 142 MVDGALPPEINARMYAGPSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 201

QY 61 SSAGLMAVAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
DB 202 SSAGLMAVAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
DB 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAVAEASDTAAANQLMNNVPOALQOAOPTGTTSSKLGGLWKTVPSPHRSPI 240
DB 322 LLEQAAVAEASDTAAANQLMNNVPOALQOAOPTGTTSSKLGGLWKTVPSPHRSPI 381
QY 241 MVSMMNNHMTNSGVSMNTLSMLKGFAPAAAQAVOTAAQNGVRAVSSILGSSSG 300
DB 382 MVSMMNNHMTNSGVSMNTLSMLKGFAPAAAQAVOTAAQNGVRAVSSILGSSSG 441
QY 301 LGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMGLGLPV 360
DB 442 LGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMGLGLPV 501
QY 361 GOMCARAGGSLGVLRVPPRPVPHSPAG 391
DB 502 GOMCARAGGSLGVLRVPPRPVPHSPAG 532

RESULT 6

US-08-818-112-111
; Sequence 111, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-112-111

Query Match

84.8%; Score 1652.5; DB 3; Length 396;

Best Local Similarity 84.9%; Pred. No. 1.3e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 1 MYDFGALPPEINSARMYAGFGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGWSWIG 60
DB 1 VVDFGALPPEINSARMYAGFGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGWSWIG 60
QY 61 SSAGLMAVAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
DB 61 SSAGLMAVAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
QY 181 LLEQAAVAEASDTAAANQLMNNVPOALQOAOPTGTTSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAVAEASDTAAANQLMNNVPOALQOAOPTGTTSSKLGGLWKTVPSPHRSPI 240
QY 241 MVSMMNNHMTNSGVSMNTLSMLKGFAPAAAQAVOTAAQNGVRAVSSILGSSSG 296
DB 241 IVSMLNNHMTNSGVSMNTLSMLKGFAPAAAQAVOTAAQNGVRAVSSILGSSSG 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMGL 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMGL 359
QY 357 GLPYGQMGARAG--GGLSGVLRVPPRPVPHSPAG 391
DB 360 GLPYGQMGARAG--GGLSGVLRVPPRPVPHSPAG 396

RESULT 7

US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid

```
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-819-111-106

Query Match      84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.3e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTGVSWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VVDFGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTGVSWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 SSAGLMVAASAPYVAMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SSAGLMVAASAPYVAMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LIATNLGQNTPAIATVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LIATNLGQNTPAIATVNEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 LLEQAAVEEASDTAAANQNMNVPOALQQAQPTGTTSSKLGGLWKTVPSPHSPISN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 LLEQAAVEEALDTAAANQNMNVPOALQQAQPTKSIWPFQDQSELWKAI SPHLSPLSN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAVMS---LGSSL 296
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 IVSMLNNHVSMTNSGVSMNTLSSMLKGFAP-AAAQAVETAQNGVQVAMSSLSGSLGSSL 299
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGPQOMLG 356
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAQAOTAPGHMLG 359
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 357 GLPVGQMGARAG--GGLSGVLVPRPPYVMPHSPAAAG 391
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 360 GLPLQLTNSGGFGGVSNALRMPRAYVMPRVPAAG 396
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
US-09-056-556-111
; Sequence 111, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids

TREATM
```

```
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-111

Query Match      84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.3e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTGVSWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VVDFGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTGVSWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 SSAGLMVAASAPYVAMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SSAGLMVAASAPYVAMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LIATNLGQNTPAIATVNEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LIATNLGQNTPAIATVNEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 LLEQAAVEEASDTAAANQNMNVPOALQQAQPTGTTSSKLGGLWKTVPSPHSPISN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 LLEQAAVEEALDTAAANQNMNVPOALQQAQPTKSIWPFQDQSELWKAI SPHLSPLSN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAVMS---LGSSL 296
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 IVSMLNNHVSMTNSGVSMNTLSSMLKGFAP-AAAQAVETAQNGVQVAMSSLSGSLGSSL 299
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGPQOMLG 356
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAQAOTAPGHMLG 359
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 357 GLPVGQMGARAG--GGLSGVLVPRPPYVMPHSPAAAG 391
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 360 GLPLQLTNSGGFGGVSNALRMPRAYVMPRVPAAG 396
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
US-09-072-596-106
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
```

REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-106

Query Match 84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.3e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAKMWDVSASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 VUDFGALPPEINSARMYAGPGSASLVAAAKMWDVSASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMVAASPYVAMSVTAGOAEILTAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLMVAASPYVAMSVTAGOAEILTAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATALLPPEAPEMITSAGG 180
DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEAPEMITSAGG 180
QY 181 LFEQAAVBEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKGLGKWTVPSPHSPISN 240
DB 181 LFEQAAVBEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKGLGKWTVPSPHSPISN 240
QY 241 MVSMAHHNSMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAONGVRAMSS----LGSSL 296
DB 241 IVSMLNHSMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAONGVRAMSS----LGSSL 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMIG 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMIG 359
QY 357 GLPVGQMGARAG--GGLSGVLRVPPRPYVPHSPAPAG 391
DB 360 GLPLGOLNTSGGGFGGVSNALRPPRAYVYVPRVPAAG 396

RESULT 10

US-08-818-112-109
Sequence 109, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-818-112-109

Query Match 76.3%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.2e-116;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAKMWDVSASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 VUDFGALPPEINSARMYAGPGSASLVAAAKMWDVSASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMVAASPYVAMSVTAGOAEILTAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLMVAASPYVAMSVTAGOAEILTAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEAPEMITSAGG 180
DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEAPEMITSAGG 180
QY 181 LFEQAAVBEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKGLGKWTVPSPHSPISN 240
DB 181 LFEQAAVBEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKGLGKWTVPSPHSPISN 240
QY 241 MVSMAHHNSMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAONGVRAMSS----LGSSL 296
DB 241 VSSMAHHNSMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAONGVRAMSS----LGSSL 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMIG 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMIG 359

RESULT 11

US-08-818-111-104
Sequence 104, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-818-111-104

Query Match 76.3%; Score 1486.5; DB 4; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.2e-116;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVMGLTVGSGWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVMGLTVGSGWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 SSAGLMVAAASPYVAMWMTAGQAEHTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SSAGLMVAAASPYVAMWMTAGQAEHTAAQVRVAAAAYETAYRLTVPPPIAENRTELMT 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LTATNLLGQNTPAIEANQAASQWQDAEAMVGYAATAATATATALLPFEAPMTINPGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 LLEQAAVVEASDTAAANQLMNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHRSPIIN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 LLEQAVAVEEADTAAANQLMNVPQALQQAQPAQGVVPSKLGGLWTAVSPHLSPLSN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 MVSMANNHMTNSGVSMNTLSMLKGFAPAAAQAVCTAAQNGVRAMSS----LGSSL 296
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 VSSIANNHSMGTGVSMTNTHLSMLKGLAP-AAQAVETAENGWAMSSLGSLGSSSL 299
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOMAAANQAVTPAARALPLTSLTSAERGGQMLG 356
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPPAWAANQAVTPAARALPLTSLTSAQAQTPAGHMLG 359
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 12
US-09-056-556-109
; Sequence 109, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; CORRESPONDENCE ADDRESS: 241
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
```

TREATM

```
;
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-056-556-109

Query Match 76.3%; Score 1486.5; DB 4; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.2e-116;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVMGLTVGSGWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVMGLTVGSGWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 SSAGLMVAAASPYVAMWMTAGQAEHTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SSAGLMVAAASPYVAMWMTAGQAEHTAAQVRVAAAAYETAYRLTVPPPIAENRTELMT 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LTATNLLGQNTPAIEANQAASQWQDAEAMVGYAATAATATATALLPFEAPMTINPGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 LLEQAAVVEASDTAAANQLMNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHRSPIIN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 LLEQAVAVEEADTAAANQLMNVPQALQQAQPAQGVVPSKLGGLWTAVSPHLSPLSN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 MVSMANNHMTNSGVSMNTLSMLKGFAPAAAQAVCTAAQNGVRAMSS----LGSSL 296
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 VSSIANNHSMGTGVSMTNTHLSMLKGLAP-AAQAVETAENGWAMSSLGSLGSSSL 299
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOMAAANQAVTPAARALPLTSLTSAERGGQMLG 356
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPPAWAANQAVTPAARALPLTSLTSAQAQTPAGHMLG 359
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
US-09-072-596-104
; Sequence 104, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; CORRESPONDENCE ADDRESS: 350
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```



```
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-111-92

Query Match          60.9%; Score 1187; DB 4; Length 263;
Best Local Similarity 99.6%; Pred. No. 3.9e-91;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 74 VAMSVTAGQAEHTAAQVRVAAAYETAYGLTVPVPIAENRAELMILLIATNLIGQNTPA 133
Db 1 VAMSVTAGQAEHTAAQVRVAAAYETAYGLTVPVPIAENRAELMILLIATNLIGQNTPA 60

Qy 134 IAVNEAEYGEHWADAAAFGYAAATATATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 193
Db 61 IAVNEAEYGEHWADAAAFGYAAATATATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120

Qy 194 TAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSNNHMSMTN 253
Db 121 TAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSNNHMSMTN 180

Qy 254 SGVSNMTNTLSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSSLGSSGLGGGVAANLGRAA 313
Db 181 SGVSNMTNTLSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSSLGSSGLGGGVAANLGRAA 240

Qy 314 SV 315
Db 241 SV 242
```

Search completed: November 21, 2003, 16:11:39
Job time : 9.24558 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 15.2499 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672a-26

Perfect score: 1949

Sequence: 1 MVDFGALPPEINSARMYAGP.....SGVLVRPFRPYVMEHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1949	100.0	391	12	US-10-084-843-107
2	1949	100.0	391	12	US-10-193-002-102
3	1949	100.0	391	12	US-10-088-732A-14
4	1949	100.0	596	9	US-09-287-849-26
5	1949	100.0	596	12	US-10-359-460-26
6	1949	100.0	596	12	US-10-098-732A-20
7	1949	100.0	600	9	US-09-287-849-22
8	1949	100.0	600	12	US-10-359-460-22
9	1949	100.0	729	12	US-10-098-732A-18
10	1949	100.0	930	12	US-10-098-732A-65
11	1944	99.7	729	9	US-09-287-849-2
12	1944	99.7	729	12	US-10-359-460-2
13	1944	99.7	729	12	US-10-098-732A-16
14	1652.5	84.8	396	12	US-10-084-843-111
15	1652.5	84.8	396	12	US-10-193-002-106

16	1486.5	76.3	359	12	US-10-084-843-109
17	1486.5	76.3	359	12	US-10-193-002-104
18	1187.5	60.9	358	9	US-09-287-849-8
19	1187.5	60.9	358	12	US-10-359-460-8
20	1187	60.9	263	12	US-10-084-843-91
21	1187	60.9	263	12	US-10-193-002-92
22	1187	60.9	263	12	US-10-098-732A-12
23	766.5	39.3	400	9	US-09-073-009-126
24	766.5	39.3	400	9	US-09-793-306-126
25	731	37.5	421	16	US-10-080-170-146
26	604	31.0	710	9	US-09-287-849-16
27	604	31.0	710	12	US-10-359-460-16
28	604	31.0	710	12	US-10-098-732A-49
29	604	31.0	856	9	US-09-287-849-12
30	604	31.0	856	12	US-10-359-460-12
31	603	30.9	423	9	US-09-073-009-142
32	603	30.9	423	9	US-09-793-306-142
33	603	30.9	423	12	US-10-098-732A-31
34	588	30.2	394	10	US-09-712-363-205
35	439.5	22.6	655	10	US-09-712-363-207
36	424.5	21.8	943	10	US-09-996-634-131
37	424.5	21.8	943	11	US-09-997-182-131
38	424.5	21.8	943	11	US-09-997-181-131
39	424	21.8	141	9	US-09-073-009-15
40	424	21.8	141	9	US-09-023-588-15
41	424	21.8	141	9	US-09-793-306-15
42	413	21.2	597	9	US-09-793-306-146
43	364.5	18.7	408	16	US-10-080-170-57
44	317	16.3	371	9	US-09-791-171-92
45	317	16.3	371	12	US-09-804-980-92

ALIGNMENTS

RESULT 1

US-10-084-843-107
; Sequence 107, Application US/10084943
; Publication No. US20030143243A1
; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/084,843

FILING DATE: 25-Feb-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967

FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.


```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-26

Query Match      100.0%; Score 1949; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSGARMYAGGSAASLVAAQWMDVSAADLFSASAQSQVVMGLTVGSMIG 60
DB 9 MVDFGALPPEINSGARMYAGGSAASLVAAQWMDVSAADLFSASAQSQVVMGLTVGSMIG 68

QY 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 69 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPVPIAENRAELMI 128

QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188

QY 181 LLEQAAAVEASDPTAAANQLMNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPISN 240
DB 189 LLEQAAAVEASDPTAAANQLMNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPISN 248

QY 241 MVSWMNNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAARALPLTSLTSAERPGQMLGGLPV 360
DB 249 MVSWMNNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAARALPLTSLTSAERPGQMLGGLPV 368

QY 301 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTAAARALPLTSLTSAERPGQMLGGLPV 360
DB 309 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTAAARALPLTSLTSAERPGQMLGGLPV 368

QY 361 GQMGARAGGSLGVLRVPRPYVMPHSPAAG 391
DB 369 GQMGARAGGSLGVLRVPRPYVMPHSPAAG 399

RESULT 6
US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 20
; TYPE: PRT
; LENGTH: 596
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (designated MTB59F)
US-10-098-732A-20

Query Match      100.0%; Score 1949; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSGARMYAGGSAASLVAAQWMDVSAADLFSASAQSQVVMGLTVGSMIG 60
DB 9 MVDFGALPPEINSGARMYAGGSAASLVAAQWMDVSAADLFSASAQSQVVMGLTVGSMIG 68

QY 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 69 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPVPIAENRAELMI 128

QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188

QY 181 LLEQAAAVEASDPTAAANQLMNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPISN 240
DB 189 LLEQAAAVEASDPTAAANQLMNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPISN 248

QY 241 MVSWMNNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAARALPLTSLTSAERPGQMLGGLPV 360
DB 249 MVSWMNNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAARALPLTSLTSAERPGQMLGGLPV 368

QY 301 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTAAARALPLTSLTSAERPGQMLGGLPV 360
DB 309 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTAAARALPLTSLTSAERPGQMLGGLPV 368

QY 361 GQMGARAGGSLGVLRVPRPYVMPHSPAAG 391
DB 369 GQMGARAGGSLGVLRVPRPYVMPHSPAAG 399

RESULT 7
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match      100.0%; Score 1949; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.2e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSGARMYAGGSAASLVAAQWMDVSAADLFSASAQSQVVMGLTVGSMIG 60
DB 9 MVDFGALPPEINSGARMYAGGSAASLVAAQWMDVSAADLFSASAQSQVVMGLTVGSMIG 68

QY 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 69 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPVPIAENRAELMI 128
```

QY 121 LIATNLGONTPTAVNEAEYGEWMAQDAAMFCYAAATATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLGONTPTAVNEAEYGEWMAQDAAMFCYAAATATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 240
Db 189 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 248
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 300
Db 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 308
QY 301 LGGGVAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
Db 309 LGGGVAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399

RESULT 8

US-10-359-460-22
; Sequence 22, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-00902005
; CURRENT APPLICATION NUMBER: US/10/359,460
; PRIOR FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/518,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-22

Query Match 100.0%; Score 1949; DB 12; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.2e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 68
QY 61 SSAGLMVAAAAPYVAMSVTAGQAEHLTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 69 SSAGLMVAAAAPYVAMSVTAGQAEHLTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
QY 121 LIATNLGONTPTAVNEAEYGEWMAQDAAMFCYAAATATATATLLPFEAPEMTSAGG 180

Db 129 LIATNLGONTPTAVNEAEYGEWMAQDAAMFCYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 240
Db 189 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 248
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 300
Db 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 308
QY 301 LGGGVAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
Db 309 LGGGVAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399

RESULT 9

US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,937
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72FMutSA
; OTHER INFORMATION: (Ra12-TbH9-Ra35MutSA)
US-10-098-732A-18

Query Match 100.0%; Score 1949; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 1.6e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 201
QY 61 SSAGLMVAAAAPYVAMSVTAGQAEHLTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 202 SSAGLMVAAAAPYVAMSVTAGQAEHLTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
QY 121 LIATNLGONTPTAVNEAEYGEWMAQDAAMFCYAAATATATATLLPFEAPEMTSAGG 180
Db 262 LIATNLGONTPTAVNEAEYGEWMAQDAAMFCYAAATATATATLLPFEAPEMTSAGG 321
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 240
Db 322 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 381
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 300
Db 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360

Db 442 LGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532

RESULT 10
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F-MAPS
; OTHER INFORMATION: (r9sf) fusion construct, TB MTB72F (Ra12-TB9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-65

Query Match 100.0%; Score 1949; DB 12; Length 930;
Best Local Similarity 100.0%; Pred. No. 2.2e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201
QY 61 SSAGLMVAASAPVYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
Db 202 SSAGLMVAASAPVYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
QY 121 LIATNLLGONTPTAI VNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 180
Db 262 LIATNLLGONTPTAI VNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAA VEEASDTAAANQLMNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
Db 322 LLEQAAA VEEASDTAAANQLMNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 381
QY 241 MVSMAHNMHMTNSGVSMNTNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSILGSSG 300
Db 382 MVSMAHNMHMTNSGVSMNTNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSILGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 442 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532

RESULT 11
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US20020009459A1

; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-2

Query Match 99.7%; Score 1944; DB 9; Length 729;
Best Local Similarity 99.7%; Pred. No. 3.9e-145;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201
QY 61 SSAGLMVAASAPVYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
Db 202 SSAGLMVAASAPVYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
QY 121 LIATNLLGONTPTAI VNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 180
Db 262 LIATNLLGONTPTAI VNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAA VEEASDTAAANQLMNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
Db 322 LLEQAAA VEEASDTAAANQLMNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 381
QY 241 MVSMAHNMHMTNSGVSMNTNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSILGSSG 300
Db 382 MVSMAHNMHMTNSGVSMNTNTLSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSILGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 442 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532

RESULT 12
US-10-359-460-2
; Sequence 2, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.

APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match 99.7%; Score 1944; DB 12; Length 729;
Best Local Similarity 99.7%; Pred. No. 3.9e-145;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAAGAFQSVVWGLTVGSMIG 60
DB 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAAGAFQSVVWGLTVGSMIG 201
QY 61 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
DB 202 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLTFEEAPEWTSAGG 180
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLTFEEAPEWTSAGG 321
QY 181 LLEQAAAVEEASDTAAANQLMNVFPQALQQLAQPTGTTTPSSKLGGLWKTVPSPHSPISN 240
DB 322 LLEQAAAVEEASDTAAANQLMNVFPQALQQLAQPTGTTTPSSKLGGLWKTVPSPHSPISN 381
QY 241 MVSMAHHMNTNSGVMTNTLSVLMKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSSG 300
DB 382 MVSMAHHMNTNSGVMTNTLSVLMKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSSG 441
QY 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQOMLGLPV 360
DB 442 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQOMLGLPV 501
QY 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
DB 502 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 532

RESULT 13
US-10-098-732A-16
Sequence 16, Application US/10098732A
Publication No. US20030175294A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guderian, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a

APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match 99.7%; Score 1944; DB 12; Length 729;
Best Local Similarity 99.7%; Pred. No. 3.9e-145;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAAGAFQSVVWGLTVGSMIG 60
DB 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAAGAFQSVVWGLTVGSMIG 201
QY 61 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
DB 202 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLTFEEAPEWTSAGG 180
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLTFEEAPEWTSAGG 321
QY 181 LLEQAAAVEEASDTAAANQLMNVFPQALQQLAQPTGTTTPSSKLGGLWKTVPSPHSPISN 240
DB 322 LLEQAAAVEEASDTAAANQLMNVFPQALQQLAQPTGTTTPSSKLGGLWKTVPSPHSPISN 381
QY 241 MVSMAHHMNTNSGVMTNTLSVLMKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSSG 300
DB 382 MVSMAHHMNTNSGVMTNTLSVLMKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSSG 441
QY 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQOMLGLPV 360
DB 442 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQOMLGLPV 501
QY 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
DB 502 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 532

RESULT 14
US-10-084-843-111
Sequence 111, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedwick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

Search completed: November 21, 2003, 16:38:13
Job time : 16.2499 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 8.15692 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-26

Perfect score: 1949

Sequence: 1 MVDFGALPPPEINSARMYAGP.....SGVLVPPRPYVMPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_76.*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1949	100.0	391	2 B70608	probable PPE prote
2	1656.5	85.0	396	2 H70741	probable PPE prote
3	1583	81.2	393	2 C70568	probable PPE prote
4	779.5	40.0	393	2 G70929	probable PPE prote
5	751	38.5	409	2 A70932	probable PPE prote
6	737.5	37.8	403	2 H70931	probable PPE prote
7	736	37.8	423	2 B70931	probable PPE prote
8	731	37.5	421	2 H87056	PPE-family protein
9	705	36.2	408	2 G70925	probable PPE prote
10	702	36.0	391	2 B70625	probable PPE prote
11	702	36.0	413	2 F70560	probable PPE prote
12	688.5	35.3	463	2 C70931	probable PPE prote
13	682.5	35.0	468	2 B70932	probable PPE prote
14	675	34.6	380	2 A70646	probable PPE prote
15	668.5	34.3	394	2 G70929	probable PPE prote
16	667.5	34.2	385	2 H70503	probable PPE prote
17	635	32.6	350	2 H70929	probable PPE prote
18	633.5	32.5	365	2 E70929	probable PPE prote
19	615.5	31.6	402	2 A70882	probable PPE prote
20	603	30.9	423	2 C70582	probable PPE prote
21	597.5	30.7	391	2 D70922	probable PPE prote
22	589	30.2	406	2 A70675	probable PPE prote
23	588	30.2	394	2 A70504	probable PPE prote
24	575.5	29.5	391	2 A70663	probable PPE prote
25	495.5	25.4	3300	2 D70575	probable PPE prote
26	473	24.3	180	2 G70834	probable PPE prote
27	452.5	23.0	3716	2 E70969	probable PPE prote
28	448.5	23.2	580	2 G70570	probable PPE prote
29	444	22.8	346	2 H70874	probable PPE prote

RESULT 1

B70608

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: B70608

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome s

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: B70608

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-391 <COL>

A;Cross-references: GB:Z93777; GB:AL123456; NID:G3261726; PIDN:CA807839.1; PID:e311073.1

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: PPE

Query Match 100.0%; Score 1949; DB 2; Length 391;

Best Local Similarity 100.0%; Pred. No. 3.3e-108;

Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSGWG 60

Db 1 MVDGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSGWG 60

QY 61 SSAGLMVAAASPYVAVMSVTAGQABLTAAQVRAAAAYETAYGLTVPPVPIAENRAELMI 120

Db 61 SSAGLMVAAASPYVAVMSVTAGQABLTAAQVRAAAAYETAYGLTVPPVPIAENRAELMI 120

QY 121 LIATNLLGNTFAIAVNEAEYGEWMAQDAANFGVAAATATATATALLPEEPAPENTSAG 180

Db 121 LIATNLLGNTFAIAVNEAEYGEWMAQDAANFGVAAATATATATALLPEEPAPENTSAG 180

QY 181 LLEQAAVVEEADTAAANQNMNVPCALQOLAQPTQGTTPSSKLGGLWKTVPSPHSPI 240

Db 181 LLEQAAVVEEADTAAANQNMNVPCALQOLAQPTQGTTPSSKLGGLWKTVPSPHSPI 240

QY 241 MYSMANNHSMNTNSGVSWTNTLSMLKGPAPAAAQAVTAQNGVRAVMSI 300

Db 241 MYSMANNHSMNTNSGVSWTNTLSMLKGPAPAAAQAVTAQNGVRAVMSI 300

QY 301 LGGGVAANIGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABRGPCQMLGGLPV 360

Db 301 LGGGVAANIGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABRGPCQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLKVPVPPRYVMPHSPAAG 391

Db 361 GQMGARAGGGLSGVLKVPVPPRYVMPHSPAAG 391

ALIGNMENTS

Db 361 GQMGRAGGGLSGVLPRPPYVPHSPAAG 391

RESULT 2

H70741
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70741
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, C.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70741
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-396 <COL>
A:Cross-references: GB:275555; GB:AL123456; NID:G3261608; PIDN:CAA999966.1; PID:e250360;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 85.0%; Score 1656.5; DB 2; Length 396;
Best Local Similarity 85.1%; Pred. No. 6.4e-91;
Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMVAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
QY 121 LIATNLGQNTPAIAVNEAEYGMWAQDAAMFGYAAATATATATLLPPEBAPMTSAGG 180
Db 121 LIATNLGQNTPAIAVNEAEYGMWAQDAAMFGYAAATATATATLLPPEBAPMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNVPOALQLOAQTOGTTTPSSKLGGLWKTVPSPHRPISN 240
Db 181 LLEQAAVEEASDTAAANQLMNVPOALQLOAQTOGTTTPSSKLGGLWKTVPSPHRPISN 240
QY 241 MVSMMNNHSMNTSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMSS----LGSSL 296
Db 241 MVSMMNNHSMNTSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMSS----LGSSL 296
QY 297 GSSGLGGVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGQMLG 356
Db 297 GSSGLGGVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGQMLG 356
QY 357 GLPVGOMGARAG--GGLSGVLPRPPYVPHSPAAG 391
Db 360 GLPLGQLTNSGGFGVSNALRNPVPRVYVPPVPAAG 396

RESULT 3

C70568
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70568
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, C.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70568
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:295390; GB:AL123456; NID:G3261766; PIDN:CAAB08702.1; PID:e316074; I
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 81.2%; Score 1583; DB 2; Length 393;
Best Local Similarity 81.5%; Pred. No. 1.4e-86;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMVAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
QY 121 LIATNLGQNTPAIAVNEAEYGMWAQDAAMFGYAAATATATATLLPPEBAPMTSAGG 180
Db 121 LIATNLGQNTPAIAVNEAEYGMWAQDAAMFGYAAATATATATALLPPEBAPMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNVPOALQLOAQTOGTTTPSSKLGGLWKTVPSPHRPISN 240
Db 181 LLEQAAVEEASDTAAANQLMNVPOALQLOAQTOGTTTPSSKLGGLWKTVPSPHRPISN 240
QY 241 MVSMMNNHSMNTSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMSS----LGSSL 296
Db 241 VSTIANNHSMNTSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMSS----LGSSL 299
QY 297 GSSGLGGVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGQMLG 356
Db 300 GSSGLGGVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGQMLG 359
QY 357 GLPVGOMGARAGGGLSGVLPRPPYVPHSPAAG 391
Db 360 GLPLGH-SVNAGSGINNLRVPRVPAVAPRTPAAG 393

RESULT 4

G70929
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, C.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN:CAAL17711.1; PID:e125460;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 40.0%; Score 779.5; DB 2; Length 393;
Best Local Similarity 43.7%; Pred. No. 4.3e-39;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 61
Db 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 62 SSAGLMVAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 121
Db 61 ASAAAEAVAPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120

```

QY 122 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFVGAATAATATATATLPPFEAPEMTSGAGGL 181
D 121 ISTNVFGONTSAIAAAEAQYGEWMAQDAAMFVGAATAATATATATLPPFEAPEMTSGAGGL 179
QY 182 LEQAAVEEASDTAAA--NOLMNNVPALQOQLAQTOGTTPSSKLGGLWK----- 229
D 180 GTQRAAVATAAGTAQSTLTETMITCLPNALQSLSPQLQSS-NGPLSLWLKILGTFENFFT 238
QY 230 -----TVSPHRPIGNVSMNNHMTNWSVGSMTNTLSSMLKGFAPAAAQAQVATAA 282
D 239 SISALLDLOFYAFFYNTTEGLPYFSIGMGNFTQSAKTL-GLIGSAAPAAVA-----AA 292
QY 283 QNGVRAMSSLSGSLGGGVAANI GRAASVGLSVPCA WAA-ANQAVTPAARALPLT 341
D 293 GDAKGLPGLGMLG-----GGPVAAGLONAAVUGKLSVFPVWGSLPGSVTPGAAFLPVS 348
QY 342 SLTSAARSGPQMLGGLPVQGMARAGGGLSGVLRVPPRPYVMPHSPAAG 391
D 349 TVSAAPAAAPGSLGGLPL-----AGAGGAGAGP-RYGFRTVMARPPFAG 393

RESULT 5
A70932
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70932
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70932
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-409 <COL>
A;Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17729.1; PID:e125461
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: PPE

Query Match 38.5%; Score 751; DB 2; Length 409;
Best Local Similarity 44.7%; Pred. No. 2.2e-37;
Matches 177; Conservative 53; Mismatches 138; Indels 28; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSGWS 61
D 1 MDFGALPPEINSARMYAGPGSGPLAAAAAANDALAAELYSAAASYSGTIEGLTVAPWMPG 60
QY 62 SAGLVAAASPYPVAMSVTAQOAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMIL 121
D 61 SSITMAAAYAPYVAVISVTAQOAEQAGQAKIAAGVYETAFATVPPVPIAENRALLMSL 120
QY 122 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFVGAATAATATATLPPFEAPEMTSGAGGL 181
D 121 VATNIFGONTFAIAATEAHYAEWMAQDAAMFVGAATAATATATLPPFEAPEMTSGAGGL 179
QY 182 LEQAAVEEASDTAAA-----NOLMNNVPALQOQLAQTOGTTPSSKLGGLWKVSPH 234
D 180 AQGAVVAQAAGAAASDITACLSQLSLIPSLQSLA--TTATATASAG--WDTV--- 232
QY 235 RSPISNNMANNHMTNWSVGS-----MTNTLSSMLKGFAPAAAQAQVATAAQNQVYRAM 289
D 233 LQSITITLANLTGYSIIGLGAIPCGWMLTFGQILGLAQNAPGVALLGPKAAAGALSPL 292
QY 290 SSL-GSSLSGS-SGLGGGVAANLGRASVGSLSVQAWAANAQVTPAARALPITS TSA- 346
D 293 APLRGYIGDITPLGGGATGCIARIAYVGLSVLPQGWAEAPVVRVAVSLPGTGAAPAL 352
QY 347 AERFGQMLGGLPVQGMARAGGGL-----SGVLRV 377
```

```

D 353 AAARPCALFGEMALSSLAGRALAGTAVRSGAARV 388

RESULT 6
B70931
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70931
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70931
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-403 <COL>
A;Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17728.1; PID:e125461
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: PPE

Query Match 37.8%; Score 737.5; DB 2; Length 403;
Best Local Similarity 42.6%; Pred. No. 1.3e-36;
Matches 181; Conservative 60; Mismatches 123; Indels 61; Gaps 12;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSGWS 61
D 5 LDFATLPPPEINSARMYAGSAGAPMLAAASAWHGLSAELRASALSYSVSLTTLTGEWHGP 64
QY 62 SAGLVAAASPYPVAMSVTAQOAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMIL 121
D 65 ASASMTAAAPYVAMSVTAQOAEQAGQAAAAAYFAAFATVPPVPIAENRAELMIL 124
QY 122 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFVGAATAATATATLPPFEAPEMTSGAGGL 181
D 125 IATNVLQGNAPAIATEAQYAEWMSQDAWAMVGYAGASNAAT-QLTPTPTVQTINAGSL 183
QY 182 LEQAAVEEASDTAAA-----NOLMNNVPALQOQLAQTOGT-----TPSS 222
D 184 AAQSAIAIAHATGASAGAOQTTLSQLIAATPSVLQGLSSSTATFASGPSGLLIGVSGSS 243
QY 223 KLGLGKTVSPHRSPISNMVMANNHMTNWSVGSMTNLS--SMLKGFAPAAAQAQV 278
D 244 WLDKWLWLLDPN-----SNFWNTIASSGLFDPNTIAPFLGLLGGVAAAADAAGDV 293
QY 279 QTAAGQNGVRAMSSLSGSL-----GSSGLGGVAAANLGRASVGSLSVQAWAANAQVATP 333
D 294 LGEATSG-----GLGGALVAPLGSAGGLGCTVAAGLGNAAATVGTLSVPPSWTAAAPLASP 348
QY 334 AAPAL---PLTSLTSAAREGPGQMLGGLPVQGMARAGGGLSGVLRVPP-----RPVMPH 386
D 349 LGSALGGTTPWAPPAPVAAAG-----MPGMPFGTMGGGFG-----RAVPQYGFRTFVAR 398
QY 387 SPAAG 391
D 399 PPAAG 403

RESULT 7
B70931
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70931
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
```

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70931
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-423 <COL>
A:Cross-references: GB:AL020201; GB:AL123456; NID:G3250699; PIDN:CAA17722.1; PID:e125461
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PPE

Query Match 37.8%; Score 736; DB 2; Length 423;
Best Local Similarity 41.8%; Pred. No. 1.7e-36;
Matches 182; Conservative 50; Mismatches 123; Indels 80; Gaps 12;

QY 2 VDFGALPPEINSGARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWGS 61
DB 1 MFDGALPPEINSGARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWGS 59
QY 62 SAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 60 SSTSVASAAAPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 119
QY 122 IATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGGL 181
DB 120 IATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGGL 178
QY 182 LEQAAVEEASPTAAN-----QIMNVPAQLQLOAQT-----QGTTP-----S 221
DB 179 GAQSSAVAQTAATAAGNLQSAFFQLLSAVPRALQGLAFTASQASATPQWVTDLGNLS 238
QY 222 SKLGLLWKTVPSPHRSPISNVSVNANNHMTNSGVSMTNTLSSMLKGFAPAAQAQVTA 281
DB 239 TFLGG--AVTGPTTP-----GVLPSPGVPYLLGQSVL-----V 271
QY 282 AQNGVRAMSLGS-----SLGSSGLGG--VAANLGRAASVGSLS 319
DB 272 TQNGQGVSAALLGKIGKPTGALAPLAEFALHTPILSGELGGSVSAGIGRAGLVGKLS 331
QY 320 VQQAANAQVTPAARALPLTSLIS--AAERGQGMQLGLPVQGMQARAGGLSLGVLR 376
DB 332 VPQGMVVAAPFPSPAALQATRLAAAPTADTGAGALLGGMALSGLAGRAAAGSTG--- 388
QY 377 VPPRPVMPHSPAAAG 391
DB 389 ---HPIGSAAAPAVG 400

RESULT 8
H87056
PPE-family protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87056
R: Cole, S.T.; Eigimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A:Title: Massive gene decay in the leprosy bacillus
A:Reference number: A85909; MUID:21128732; PMID:11234002
A:Accession: H87056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL450380; NID:G13093150; PIDN:CAC31563.1; GSPDB:GN00147
C:Genetics:
A:Gene: MLI182

Query Match 37.5%; Score 731; DB 2; Length 421;
Best Local Similarity 41.1%; Pred. No. 3.4e-36;
Matches 174; Conservative 55; Mismatches 160; Indels 34; Gaps 8;

QY 1 MVDGALPPEINSGARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWGS 60
DB 1 MFDGALPPEINSGARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWGS 59
QY 61 SAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 120
DB 60 ESAAALAEVTPPEKWLTONAASAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 119
QY 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGGL 180
DB 120 LIMSIFGQNSTAIAEKEAEYTEMWQDAAMTSYQASVLEAVGATKATAPPLGVNEVG 179
QY 181 L-----LEQAAVEEASPTAANQIMNV-----VPAQLQ-----LAQP 214
DB 180 LAQEVVEEVVEEVVEEVVEEVEAEQAISQAALDQAVNEGMEATVPVQDQVNVDAVTP 239
QY 215 TQGTTPSSKLG--LWKTVPSPHRSPISNVSVNANNHMTNSGVSMTNTLSSMLKGFAPA 272
DB 240 QTAVPDSSSSAAAPQLNGGFAQLHLSPLINDTLNHNHAGMANAGLSLVNCGWSAMKSLAP- 298
QY 273 AAQAQVTAQNGVRAMSVSLGSSGLGGVAAANLGRAASVGSLSVQAAANQAVT 332
DB 299 TTTTAAESAFKAMGSVQSTGRGLGSSGGHVTAGLGRAASTIGSLRVPQTWTASQPTV 358
QY 333 PAARALPLTSLTSAARPGCOML--GGLPVQGM--CARAGGGLSGVLVPRPPVMPHSP 388
DB 359 AATRALSPARKAVAVATESAPLLGGGLPMPVPGGSGTGGVNTALRLQPRAFVMPRP 418
QY 389 AAG 391
DB 419 AAG 421

RESULT 9
G70925
Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70925
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70925
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-408 <COL>
A:Cross-references: GB:274024; GB:AL123456; NID:G3250700; PIDN:CAA98377.1; PID:e1301025;
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PPE

Query Match 36.2%; Score 705; DB 2; Length 408;
Best Local Similarity 41.8%; Pred. No. 1.1e-34;
Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;

QY 2 VDFGALPPEINSGARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWGS 61
DB 1 MFDGALPPEINSGARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWGS 60
QY 62 SAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 61 AAASMAVAATPVYVAVLWLSATAGQAEQAGMQARAAAAAYELAFAMTVPPVPPVIANRALLVAL 120
QY 122 IATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGGL 181
DB 121 VATNFFGONTPTAIAATEAQVAENWADAAAMVAYAGSAAIAT-ELTPFTTAAPTVTSIPAAL 179
QY 182 LEQAAA-----VEEASDTAAANQIMNV-----VP---QALQ--LAQGTGTTSSKL--- 224

Db 180 ACQAAATVSSTVPLATTAAPVQLLQQLSLSLIPWYALQQLAENLLGLTPDNRTIV 239
 QY 225 -----GGLWKTVSPPHRSPISNMWSMANNHSMNTSGVSMNTLSSMLKGFAPAAA 275
 Db 240 RLIGISYFDEGL-----LQFEASLAQAATPGTGGAG--DSGSSVLDSWGPPIFA 287
 QY 276 QAVQTAAQNGVRMSSL--GSSLSG-----SLGGGVAANLGRAASVGSLS 319
 Db 288 -----GPRASPSVAGGAVGVQTPQPYWYALDRSIGSVSAALKGSSAGSL 338
 QY 320 VPOWAAANQVTPAARALP---LTSLSAAERPGQMLGLPVQMGARAGGSLGVLR 376
 Db 339 VPPDWAARWANPAWLPGDDVTALRGTAENA---LLRFPWASAGQSTGGGF--VHK 393
 QY 377 VPPRYNMPHSPAAG 391
 Db 394 YGFRILAVNQRPFFAG 408

RESULT 10
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C/Species: Mycobacterium tuberculosis
 C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C/Accession: B70625
 R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A/Reference number: A70500; MUID:98295987; PMID:9634230
 A/Accession: B70625
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-391 <COL>
 A/Cross-references: GB:Z95436; GB:AL123456; NID:g3261714; PIDN:CAB06873.1; PID:e304546;
 A/Experimental source: strain H37RV
 C/Genetics:
 A/Gene: PPE

Query Match 36.0%; Score 702; DB 2; Length 391;
 Best Local Similarity 42.3%; Pred. No. 1.6e-34;
 Matches 166; Conservative 62; Mismatches 150; Indels 14; Gaps 7;

QY 2 VDFGALPEINSARMYAGPGSASLVAAQMWDSVASDLFSASAFQSVVWGLTVGWSIGS 61
 Db 1 MDFGALPEINSARMYAGAGAGPMMWAGAAWNGWLAELGTTAASYESVITRLTTESWMPG 60
 QY 62 SAGLMVAASPYVAMSVTAGCAELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMIL 121
 Db 61 ASMAWAAQPYLAWLITVTAEEAAHAGSOAMASAAAYEAYAMTVPEVVAANKALLAAL 120
 QY 122 IATNLLGONTTAAIAVNEAEYEMWAOQDAAMFGVAATATATATALLPPEAPENTSAGGL 181
 Db 121 VATNVLGINTTAIMATEALYAEWMAQDALAMYGVAASG--AAGMLQLPLSPSQTNPGL 179
 QY 182 LEQAAVEEASDTAAANQ-----LMNVVPOALQOLAQPTQCTTPSSKLGGLWKTVSPPHRS 236
 Db 180 AAQSRVGSRAATAAVNQVSDVLSLIPNAVSGLASPVTSVLDSTGLSGLIIDALLA 239
 QY 237 P--ISNMVSMANNHMS-MTNSGVSMNTLSSMLKGFAPAAAQAVOTAQNGVRAMSSLG 293
 Db 240 TPFVANIINSVNTAAWYVNAIPTAIFLANALNSGAPVAIEGAIEAEG--AASAAA 296
 QY 294 SSLSGSLGGVVAANLGRAASVGSLSVPOWAAANQVTPAARALPLTSLTSAERGPQG 353
 Db 297 AGLADSVTPAGLSAGLGEATLVGLSLVFAAWSTAAPTATACATALEGSGWTVAAEEA-GP 355
 QY 354 MLGGLPVQMGARAGGSLGVLRVPPRPVMP 385
 Db 356 VTGMWP-GMASAAKGTGAYAGFRYGFKPTWMP 386

RESULT 11

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C/Species: Mycobacterium tuberculosis
 C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C/Accession: F70560
 R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A/Reference number: A70500; MUID:98295987; PMID:9634230
 A/Accession: F70560
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-413 <COL>
 A/Cross-references: GB:Z95436; GB:AL123456; NID:g3261770; PIDN:CAB08826.1; PID:e316565;
 A/Experimental source: strain H37RV
 C/Genetics:
 A/Gene: PPE

Query Match 36.0%; Score 702; DB 2; Length 413;
 Best Local Similarity 39.1%; Pred. No. 1.7e-34;
 Matches 168; Conservative 61; Mismatches 145; Indels 56; Gaps 7;

QY 1 MVDFFGALPEINSARMYAGPGSASLVAAQMWDSVASDLFSASAFQSVVWGLTVGWSIG 60
 Db 1 MLDFAQLPPEVNSALMYAGPGSGPMLAAANAAEALAELOTTASIVDALITGLADGPWQ 60
 QY 61 SSAGLMVAASPYVAMSVTAGCAELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
 Db 61 SSAASWVAATPQVWLRLSTRAGCAEQAGQAASAAAYEAFATVPPEIAANRALLMA 120
 QY 121 LIATNLLGONTTAAIAVNEAEYEMWAOQDAAMFGVAATATATATALLPPEAPENTSAGS 180
 Db 121 LIATNLLGONTTAAIAVNEAEYEMWAOQDAAMFGVAATATATATALLPPEAPENTSAGS 179
 QY 181 LLEQAAVEEASDTAAANQLMNVVPOALQOLAQPTQCTTPSSKLGGLWKTVSPPHRSPI 240
 Db 180 LASQAAASVGQAVSGAANAQALTDIPKAL-----FGLSGIFTNEPPWLTDLGK 226
 QY 241 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAA---AAQAVOT----- 280
 Db 227 ALGTLGHTWSSDGSGLIVGGVLDGDFVQVGTGSAELDASVAMDTFGKWSPARLMTVQFKD 286
 QY 281 -----AAQNGVRAMSSISLSSGLSGGGVAAANLGRAASVGSLSVPOA 323
 Db 287 YFGLAHDLPKWASEGAKAAGEAKALPAAPVPAIPASGL--SGVAGAVGQAASVGGGLKVP 345
 QY 324 WAAANQVTPAARALPLTSLTSAERGPQMLGGLPVQMGARAGGSLGVLR--RVPPRP 381
 Db 346 WTATTPAASPVAALASNLGAAAAAEGSTHAFGGMPL--MGSGAGRAFNNFAAPRYGFKP 403
 QY 382 VYNPHPSPAAG 391
 Db 404 TVIAQPPAGC 413

RESULT 12

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C/Species: Mycobacterium tuberculosis
 C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C/Accession: C70931
 R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: C70932

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-463 <COL>

A;Cross-references: GB:AL020201; GB:AL123456; NID:93250699; PIDN:CAAL1773.1; PID:e125462

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: PPE

Query Match 35.3%; Score 688.5; DB 2; Length 463;

Best Local Similarity 42.8%; Pred. No. 1.2e-33;

Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;

QY 2 VDFGALPPPEINARMYAGGSAASLVAAQWDSVADLPSAASAFQSVVWGLTVGSWIGS 61

DB 1 MDFGVLPPEINSGRMVAGGSGPMIAAAAADWDLATLQSTAAADYGSVISVLT-GVWSGQ 59

QY 62 SAGLWVAASPYYVAMSVTTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 121

DB 60 SSGTWAARAPYVAMSVATAALAREAAQAASAAAYEAPATVPPVVAANRAELAVL 119

QY 122 IATNLGQNTPAIVNAEAYGEMWQADAAAFYAAATATATATLTPPEEAPEMTSAGGL 181

DB 120 AATNIFGQNTGAIAAEARYAENWQADAAAFYAAATATATATLTPPEEAPEMTSAGGL 178

QY 182 LEQAAVEEASDTAAANQNMNVPOALQAOPTQGTTPSSKLGGLKWTVS--PHRSPI- 238

DB 179 ATQGVAAQVAGSAGN-ARSLVSEVLEFLA--TAGTNYKTVASLMNAVGVFPASSVY 235

QY 239 -----SNMVSNNHMTNSGVSMNTLSSMLKGFAFAAAQAQVATAAQNQVRA 288

DB 236 NSMLGLGFAESKNVLPANDTVISTFGVQFKFNPVTPFEDLIPK----- 283

QY 289 MSLSGLSGLG-----SSGLGG--GVANLGRAASVGSLSVPOAWAANQAVTPAARALPL 340

DB 284 -SALGAGLGLRGAISGLGSTAPASAGASQAGSVGMSVPPSWAAATPAIRTVAAVPSS 342

QY 341 TSITS--AARPGQML-----GGLPVGQMGARAGGLSGVLRV 377

DB 343 TGLQAVPAAISEGSLLSQVALASVAGGALGGAARATGGFLGGGRV 389

RESULT 13

B70932

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: B70932

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: B70932

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-468 <COL>

A;Cross-references: GB:AL020201; GB:AL123456; NID:93250699; PIDN:CAAL1773.1; PID:e125462

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: PPE

Query Match 35.0%; Score 682.5; DB 2; Length 468;

Best Local Similarity 40.9%; Pred. No. 2.8e-33;

Matches 164; Conservative 59; Mismatches 125; Indels 53; Gaps 9;

QY 2 VDFGALPPPEINARMYAGGSAASLVAAQWDSVADLPSAASAFQSVVWGLTVGSWIGS 61

DB 1 MDFGLQFFETTSITGEMYLGPAGFMIAAAVAVDWDLAELAQSMMAASVASIVEGSAESWLGP 60

Db 290 CAASVGNIVLASGRANISIGLSVPPGSWAAPSTRFVSAISPAGLTTLTLPFGTDVAEHGMFG- 348

Qy 354 MLGGLPYQGQARAGGLSGVL RVPPRPYVMHPSPAAG 391

Db 349 -VPGVPV-----AAGRASGVLPYGVRLTVMAGHPSPAAG 380

RESULT 15
G70861

Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70881

R: Cole, S. T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: G70881

A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-394 <COL>
A: Cross-references: GB:AL008967; GB:AL123456; NID:G3261491; PIDN:CAAL1564.1; PID:ell17389
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: PPE

[illegible]

Search completed: November 21, 2003, 16:09:54
Job time : 9.15692 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 4.61043 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-26

Perfect score: 1949
Sequence: 1 MVDFGALPPINSARMYAGP.....SGVLRVPPRYVMPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1656.5	85.0	396	1 YD61_MYCTU	Q11031 mycobacteri
2	705	36.2	408	1 Y992_MYCTU	Q10813 mycobacteri
3	688.5	35.3	463	1 Y102_MYCTU	O53951 mycobacteri
4	444	22.8	487	1 Y442_MYCTU	P42611 mycobacteri
5	426.5	21.9	443	1 Y878_MYCTU	Q10540 mycobacteri
6	418	21.4	678	1 Y248_MYCTU	Q10778 mycobacteri
7	364.5	18.7	408	1 SRA_MYCLE	Q07297 mycobacteri
8	324.5	16.6	463	1 Y096_MYCTU	Q10892 mycobacteri
9	324	16.6	434	1 YU18_MYCTU	P31500 mycobacteri
10	321.5	16.5	435	1 YU21_MYCTU	O53268 mycobacteri
11	233.5	12.0	178	1 YV29_MYCTU	O06246 mycobacteri
12	217.5	11.2	176	1 YV25_MYCTU	O50703 mycobacteri
13	153.5	7.9	860	1 ELS_MOUSE	P54320 mus musculu
14	151.5	7.8	232	1 Y126_MYCTU	Q50702 mycobacteri
15	150	7.7	864	1 ELS_RAT	Q99372 rattus norv
16	143.5	7.4	730	1 ELS_HUMAN	P15502 homo sapien
17	139.5	7.2	881	1 PRY3_YEAST	P47033 saccharomyc
18	135	6.9	907	1 A180_HUMAN	O50641 homo sapien
19	133.5	6.8	825	1 ICP0_HSV2H	P28284 herpesimp
20	132	6.8	1120	1 STFR_ECOLI	P76072 escherichia
21	132	6.8	2090	1 N214_HUMAN	P35658 homo sapien
22	131	6.7	836	1 Y236_BPMLS	Q05233 mycobacteri
23	131	6.7	1150	1 APNU_PIG	P12021 sus scrofa
24	129	6.6	1783	1 RAA3_CHLRE	Q9fec4 chlamydomon
25	128	6.6	779	1 SRP_DROME	P52172 drosophila
26	127.5	6.5	790	1 ANP_NOTCO	P24856 notochenia
27	126	6.5	354	1 YAU6_SCHPO	Q10169 schizosacch
28	125.5	6.4	1211	1 BUN2_DROME	Q24523 drosophila
29	125	6.4	577	1 CST2_HUMAN	P33240 homo sapien
30	124	6.4	2090	1 HFC1_MESAU	P51611 mesocricetu
31	123.5	6.3	677	1 Y136_MYCTU	Q50597 mycobacteri
32	123	6.3	432	1 YF10_MYCTU	P71789 mycobacteri
33	123	6.3	1140	1 YN96_YEAST	Q44893 saccharomyc

ALIGNMENTS

RESULT 1

YD61_MYCTU STANDARD; PRT; 396 AA.

AC Q11031;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein Rv1361c.
GN Rv1361c OR MT1406 OR MICF02B10.25C.

OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean M.A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z75555; CAA99986.1; -.
CC DR EMBL; AE007013; AAA45669.1; -.
CC PIR; H70741; H70741.
CC TIGR; MT1406; -.
CC TubercuList; Rv1361c; -.
CC InterPro; IPR000030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.

34 122 6.3 394 1 HYF1_ALCEU
35 121.5 6.2 1025 1 SLAP_CAUCR
36 121.5 6.2 1845 1 Z236_HUMAN
37 121 6.2 1199 1 P121_RAT
38 119.5 6.1 635 1 HMLA_DROME
39 119 6.1 915 1 A180_RAT
40 118.5 6.1 580 1 EXPR_XANCP
41 118.5 6.1 2038 1 FSH_DROME
42 118 6.1 1508 1 BCSC_XANAC
43 117 6.0 444 1 Y808_CHLPN
44 117 6.0 652 1 PICA_HUMAN
45 117 6.0 774 1 STP_LAMBD
P45805 alcaligenes
P35828 caulobacter
Q9u136 homo sapien
P52391 rattus norv
P10105 drosophila
Q05140 rattus norv
P23314 xanthomonas
P13709 drosophila
P58938 xanthomonas
Q92798 chlamydia p
Q13492 homo sapien
P03764 bacterioph


```

Db      284 -SALGAGLGRSAISSGLGSTAPAIASAGASQAGSVGMSVPPSWAATPAIRTVAAVFSS 342
QY      341 TSLTSS-AAERPGQML-----GGLPVGQMGARAGGGLSGVLRV 377
Db      343 TGLQVPAARAISEGSLSSQVALASVAGGALGGAAARATGFLGGRV 389

RESULT 4
Y442 MYCTU
ID Y442 MYCTU STANDARD; PRT; 487 AA.
AC P426J1; O53727;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV0442c.
GN RV0442C OR MT0458 OR MTV037.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterinae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Erdmann;
RX MEDLINE=87137260; PubMed=3029018;
RA Shinnick T.M.;
RT "The 65-kilodalton antigen of Mycobacterium tuberculosis.";
RL J. Bacteriol. 169:1080-1088 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Englemer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G. (1998).
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551; Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Hart D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M15467; AAA88235.1; ALT INIT.
CC EMBL; AL021932; CAAL7399.1; --
CC EMBL; AS006548; AAK44681.1; --
CC PIR; C70830; C70830.
CC TIGR; MT0458; -.
CC TubercuList; RV0442c; -.
CC InterPro; IPR000030; Microbac_PPE.
CC InterPro; IPR002989; Mycobac_pentapep.
CC Pfam; PF01469; Pentapeptide_2; 5.

```



```
RESULT 6
VF48 MYCTU
ID_YF48 MYCTU STANDARD; PRT; 678 AA.
AC Q10778;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV1548c.
GN RV1548C OR MT1599 OR MTCY48.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
PC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayan L.A., Khouri M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
DR EMBL; Z74020; CAA98335.1; -
DR EMBL; AE007026; BAK45866.1; ALT_INIT.
DR FIR; A70762; A70762.
DR TIGR; MT1599; -.
DR Tuberculist; RV1548c; -.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 11.
DR Pfam; PF08823; PPE; 1.
DR Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34
FT POTENTIAL.
FT TRANSFAM 180 200
FT CONFLICT 258 258 D -> G (IN REF. 2).
SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;

Query Match 21.4%; Score 418; DB 1; Length 678;
Best Local Similarity 31.9%; Pred. No. 2.9e-17;
Matches 106; Conservative 48; Mismatches 140; Indels 38; Gaps 7;

QY 2 VDFGALPEINSARMYAGGSGASLVAAQWDSVASDLFSAASQSVVWGLTVGWSIGS 61
DQ 1 MNFSLVLPPEINSALNFAGAGGCPMLAASANTGLAGDLGSAASFSATVTSQATGSGWGP 60
DQ 62 SAGLMVAASPVAMSVTAGQAEILTAQAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 121
```

```

CC -----
DR EMBL; U00015; AAC43220.1; -
DR EMBL; X68431; CAA48480.1; -
DR EMBL; Z21952; CAA79950.1; -
DR EMBL; Z97179; CAB09938.1; -
DR EMBL; AL583918; CAC29919.1; -
DR PIR; C86960; C86960.
DR PIR; S33522; S33522.
DR PIR; S39872; S39872.
DR Leproma; ML0411; -
DR InterPro; IPR000030; Microbac_PPE.
DR Antigen; Repeat; Complete proteome.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 209 235 2 X 6 AA REPEATS OF S-V-A-Q-S-E.
FT REPEAT 209 214 1.
FT REPEAT 230 235 2.
FT CONFLICT 132 132 T -> S (IN REF. 2).
FT CONFLICT 189 189 S -> L (IN REF. 2).
FT CONFLICT 191 191 H -> D (IN REF. 2).
FT CONFLICT 292 292 P -> L (IN REF. 2).
SQ SEQUENCE 408 AA; 42466 MW; 5C0C2BE0D6E6A9D8 CRC64;

Query Match
Best Local Similarity 18.7%; Score 364.5; DB 1; Length 408;
Matches 115; Conservative 72; Mismatches 163; Indels 97; Gaps 11;

QY 1 MVDFGALPPEINARMYAGFGSASLVAAQAQMDSDVASFSAFQSVVWGLTVGSGWIG 60
Db 1 MEDEMYVSEVNAFLMRGPGSTPLWGAEMISLAEQLMEAAQESVDIVVAVPASTAG 60
QY 61 SSAGLVAAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 ETSMDLASRSTVFVAVLDGNAENAGLIARVLHAYAFEPERAGMVPLTLVLGNIHTMA 120
QY 121 LIATNLGONTPIAIVNEAYEGEMVAQDAAMFGYAAATATATATLPEEAPEMTSAGG 180
Db 121 LXAINFGQVSTTVALEADYDLWQVNSTAMTTRDVTLRGTGXENFEPAPQLVSR-Y 179
QY 181 LLEQAAVBEASDTAAANQMNVPALQQLAQ-
Db 181 CMDRDSVNSFSSSSSDSLYSIDNLYDVAQSEHSGSDMSQSVNTCGSVAQSELCD 239
QY 214 ----PTQ-----GTFPSKGLGLWKTVPHRSPISNNVSMANNHMTNMTSGVSMNT 261
Db 240 PFGTSPQSQSQNDLSATSITQQLGL-----DSIISASASLITNNS--ISSST 286
QY 262 LGSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSGLGQVAAAN-----LGRAASVG 316
Db 287 ASSIM----PIVASQVTEILGRSQV-AVEKMTQISSTAVSDVAASKVWAGVQAVSVG 341
QY 317 SLISVPOAANAQVTPARALP--LTSLSAERPGQMLGGLPVQGMGARAGGLSGV 374
Db 342 ALRVPENWATASQPVNATASHVPAGCSAITTA-----VSGPLEGV 381
QY 375 LRVPFRPYVMPHSPAG 391
Db 382 TQ--PAEEVLITASVAGG 396

RESULT 8
Y096_MYCTU STANDARD; PRT; 463 AA.
AC Q10892;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV0096.
GN RV0096 OR MT0105 OR MTCY251.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;

```

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74410; CAA8932.1; -
DR EMBL; AE006922; AAK4327.1; -
DR TIGR; H70750; H70750.
DR TIGR; MT0105; -
DR TubercuList; RV0096; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 276 295 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
SQ SEQUENCE 463 AA; 46894 MW; 42D9D6GA03D0DD8 CRC64;

Query Match
Best Local Similarity 16.6%; Score 324.5; DB 1; Length 463;
Matches 112; Conservative 54; Mismatches 144; Indels 93; Gaps 12;

QY 6 ALPPEINARMYAGPGSASLVAAQAQWDSVASFSAFQSVVWGLTVGSGWIGSAGL 65
Db 2 AIPPEVHSGLLSAGCGPGSLLVAAQWQELSDQYALACELGQLLGEVQASSWQGTAAQ 61
QY 66 MVAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPVIAENRAELMIATN 125
Db 62 YVAAHGFLANLEQTAINSAVTAQHVAAAAYCSLAAMPTPAELAAHAIHGVLIATN 121
QY 126 LQONTPTAIVNEAYEGEMVAQDAAMFGYAAATATATATLPEEAPEMTSAGGLLEQA 185
Db 122 FFGINTVPIALNEADYVRWMLQAADTMAAYQAVADAATVAVPSTQPPAPPGG---- 176
QY 186 AAVEEASDT-----AAANQIMNNVPALQQLAQPTQ-----GTTFSSKL----- 224

```



```
Db 177 -----DAADTRLDVLSIGQLIRDI-----LDFIANPYKYFEFFQGFSPAVTVVLAVAL 229
Qy 225 -----GGIM-----KTVSPHRSPTSNMVMNHNMTSGVSNMTLNSMLK-----GF 269
Db 230 QLYDFLWPPYASYGLILLPFTF-----TLSSALTALSALHLLNLPAGL 275
Qy 270 APAAAQAQVQTAQAQNGVRAMSSGLSSGLGGVAAANLGRAASVGLSVLPQAWAANQ 329
Db 276 LPFAAA-----LPGDOWGANLAVATPATAAF-----CGSP 308
Qy 330 AVTPAARALPLTSLTSAERPG-----QMLGLPLVG-QMGARAG 368
Db 309 PTSNPAPAPASNSVGSAAAPGISYAVPGLAPPGVSSGPAG 351

RESULT 9
YU21_MYCTU
ID YU21_MYCTU STANDARD; PRT; 434 AA.
AC P31500; O53265;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical PFE-family protein RV3021c.
GN RV3018C OR MT3098/MT3101 OR MTV012.32C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Taylor K., Whitehead S.,
RA Sulston J.E., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1531 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 150-374 FROM N.A.
RC STRAIN=Isolate 50410;
RA Patki A.H., Dale J.W.;
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -!- CAUTION: In strain Oshkosh the gene for this protein is
CC interrupted in position 307 by an 186110 element
CC -!- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
CC reductase.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 294; 337 and 355.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; AL021287; CAA16103.1; -
DR EMBL; AE007129; AAK47427.1; ALT SEQ.
DR EMBL; AE007129; AAK47430.1; ALT SEQ.
DR EMBL; X59271; CAA41961.1; ALT_FRAME.
DR PIR; E70857; E70857.
DR TIGR; MT3098; -
DR TIGR; MT3101; -
DR TubercuList; RV3018c; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;

Query Match 16.8%; Score 324; DB 1; Length 434;
Best Local Similarity 28.0%; Pred. No. 4.7e-12;
Matches 115; Conservative 56; Mismatches 178; Indels 62; Gaps 12;

Qy 6 ALPPEINSARMYAGPGSASIVAAAQWDSVASDLFSAASAFQSVWGLTVGSGWIGSAGL 65
Db 8 ASPPEVHSALLSAGPGPGSLQAAAAGWSALSIEYAAVAQELSVVVAAGVWQGSDEL 67
Qy 66 MVAASPYVAMSVTAGQAEELTAAQVRVAAAYETAYGLTVPPVIAENRAELMILIATN 125
Db 68 FVAAYVYVAVLVOAGSADGAAAAAGEHEAAAAGVVCALAEPMPTLPFLAANHLTHAVLATN 127
Qy 126 LLQONTPTAIVNEAEYCEMNAODAAAFVGAATATATATLTPFEAPENMTSAGGLLEQA 185
Db 128 FFGINTIPALNEADYVRMWQATVNSAYEAVVGAALVATPHGTGPAPVIVKEG----- 181
Qy 186 AAVEEASDSTAAN-----QLMNVVPAQLQLAQTOCTTPSSKLGLLWKTVSPH 234
Db 182 -ANEASNAVAAATITPFPHWEIVQLEETFAAYDQVLSALLSELPA--VAWVWFOLFVD 237
Qy 235 ---RSFSINMVMNHNMTSGVSNMTLNSMLKGP----- 270
Db 238 ILGNFIIGFTITLASNAQLTFEAINASVAVGLYAIAGVIDIVVEWIGNLFGVYVPLL 297
Qy 271 -----PAAAQAQVQTAQAQNGVRAMSSGLSSGLGGVAAANLGRAASV-GSLSVPOA 323
Db 298 GGPLLGAALAAVVPVGVAGLAVAGLAAL-PAVGA--AGAPAAVGVSAFVSGVSVPOA 354
Qy 324 WAAANQAVTPAARALPLTSLTSAERPGQMLGGLPVQNGARAGGGLSGV 374
Db 355 RLVS--AVEPAPASTSVSVL--ASDRGAGAL--GF-VGTAGKESVGQAGL 398

RESULT 10
YU21_MYCTU
ID YU21_MYCTU STANDARD; PRT; 435 AA.
AC O53268; O53269;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PFE-family protein RV3021c/RV3022c.
GN RV3021C/RV3022C OR MT3106 OR MTV012.35C/MTV012.36C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Taylor K., Whitehead S.,
RA Sulston J.E., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1531 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 150-374 FROM N.A.
RC STRAIN=Isolate 50410;
RA Patki A.H., Dale J.W.;
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -!- CAUTION: In strain Oshkosh the gene for this protein is
CC interrupted in position 307 by an 186110 element
CC -!- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
CC reductase.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 294; 337 and 355.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

"Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
Nature 393:537-544 (1998).
[2]

RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a frameshift in position 82.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021287; CAA16106.1; ALT_FRAME.
DR EMBL; AL021287; CAA16107.1; ALT_FRAME.
DR EMBL; AB007129; AAK47435.1; -.
DR TIGR; MT3106; -.
DR TubercuList; Rv3021c; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 299 299 G -> A (IN REF. 2).
FT CONFLICT 317 320 LAGV -> VTGL (IN REF. 2).
FT CONFLICT 326 326 L -> V (IN REF. 2).
SQ SEQUENCE 435 AA; 42876 MW; 3B157643EAA8484A CRC64;

Query Match 16.5%; Score 321.5; DB 1; Length 435;
Best Local Similarity 26.5%; Pred. No. 6.6e-12;
Matches 115; Conservative 53; Mismatches 159; Indels 107; Gaps 12;
QY 6 ALPPEINARMVAGPGSASLVAAQWDSVADLFAASAFQSVVWGLTVGWSAGL 65
DB 8 ASPPEVHALLSAGPGSLQAAAGWSALSAYAAVAGLSVVVAAGVAGVQGPSREL 67
QY 66 MVAASPYVAMSVTAGQELTAAQVRVAAAYETAYGLTVPPPPVIAENRAELMILITN 125
DB 68 FVAAYPYVAVLWVQASADSAAGAAAGEHEAAAGVVCALAEPTLPBLAAHMLTHAVLVATN 127
QY 126 LLQONTPTAIVNEABYGEVWQDAAAMFCVAA----- 157
DB 128 FFGINTIPIALNEADTVRMVQATVMSYAEVVGAAVLVATHTGTPAPVIVKPGANEASN 187
QY 158 ATATATATALLPPE-----APEMTSAGGLEQAAYVEEASDTAAANQLMNNYFQALQ 209
DB 188 AVAAATITPPFGELAKFLEMAAQAPTEVGLIMKSAEAWGVFELITGLVNFEP---- 243
QY 210 QLAQPTQGTTPSKLGLWKTVPSPHSPISNVSM-----ANNH 248
DB 244 -----LVV-----LTGMIDMFFATVGFALGVFLVPLLEFAVVLE 278
QY 249 MSMNTSGVSWNTLSS-----MLKGFAPAAAAQAVQTAQNGVRAMSSLSGSSGL 301
DB 279 LAITSIGLIISNIFGAIPVLGPGLLGALAAAVVPGVAGLVAGLAALPAVGAAGAP-- 336
QY 302 GGGVAANLGRASV-CSLSPQAWAANAQVTPAARALPLTSLTAAERGPQMGLGLPV 360
DB 337 -----AALVGSVAPVSGGVSPQARLVS--AVEPAPASTSVSL--ASDRGAL--GF-V 385
QY 361 GQMGARAGGGLSGV 374

Db 386 GTAGKESVGGQFAGL 399

RESULT 11

YY29_MYCTU STANDARD; PRT; 178 AA.
ID YY29_MYCTU
AC O06246;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein Rv3429.
GN Rv3429 OR MT3533 OR MTCY77.01.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95389; CAB08678.1; -.
DR EMBL; AE007158; AAK47873.1; -.
DR TIGR; C70975; C70975.
DR TubercuList; Rv3429; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 178 AA; 19811 MW; 8BE1FC025ABFBEA6 CRC64;
Query Match 12.0%; Score 233.5; DB 1; Length 178;
Best Local Similarity 35.9%; Pred. No. 3e-07;
Matches 65; Conservative 25; Mismatches 76; Indels 15; Gaps 3;
QY 7 LPPEINARMVAGPGSASLVAAQ-----MWDs---VASDLFSAASAFQSVWGLTVGWSI 59
DB 5 IPPEYISNIIYEGPGADUSAAAEQRLMYNSANMTAKSLTDLRLGELQ-----NWK 56
QY 60 GSSAGLMMVAASPYVAMSVTAGQELTAAQVRVAAAYETAYGLTVPPPPVIAENRAELM 119

Db 57 GSSDLMAAGRYLDMTKHSRQILETAYVIDELAVYVEETRHKVVPATIANNREEVH 116
 QY 120 ILIATNLGQNTPTALVNEAEVGEWMAODRAAMFGYAAATATATATLLPFEAPMTSAG 179
 Db 117 RLIASNVAGVNTPALAGLDAQYQYTRAQNTIAVMNDYQSTARFILAYLPRWQEPPIYGGG 176
 QY 180 G 180
 Db 177 G 177

RESULT 12

Y125 MYCTU
 ID Y125 MYCTU STANDARD; PRT; 176 AA.
 AC Q50703;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical PFE-family protein RV3425.
 GN RV3425 OR MTCY78.04C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=96295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Bartell B.G.,
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z77165; CAB01031.1; -
 DR PIR; F70738; F70738.
 DR Tuberculid; RV3425; -
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 176 AA; 19855 MW; B8CFE2E9463B87B0 CRC64;

Query Match 11.2%; Score 217.5; DB 1; Length 176;
 Best Local Similarity 33.5%; Pred. No. 2.5e-06;
 Matches 58; Conservative 31; Mismatches 83; Indels 1; Gaps 1;
 QY 7 LPPEINARMYAGPGSASLVAAQWDSVADLSFSAASAFQSVVWGLTVGSGSAGLM 66
 Db 5 IPAEIYSNIIEYFGGADSLFFASGQRLRELAYSVETTAESLEDELD-ENWKGSSDLL 63
 QY 67 VAASPYVAVMSVTAQOELTAQVRAVAAAYETAYGLTVPPPVIAENRAELMILITNL 126
 Db 64 ADAVERYLQWLKSKHSQKHAARWINGLANYNDRRKYVPPPEETAAENREERELIASNV 123
 QY 127 LGONTPALVNEAEVGEWMAODRAAMFGYAAATATATATLLPFEAPMTSAG 179
 Db 124 AGVNTPALADLDAQDQVRRARVAVNAYVSWTRSAUSDLPFRWEPPIYRGG 176

RESULT 13

ELS_MOUSE
 ID ELS_MOUSE STANDARD; PRT; 860 AA.
 AC P54320;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Elastin precursor (Tropoelastin).
 GN ELN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Lung;
 RX MEDLINE=95130069; PubMed=7829060;
 RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
 RT "Use of an intron polymorphism to localize the tropoelastin gene to
 RT mouse chromosome 5 in a region of linkage conservation with human
 RT chromosome 7."
 RL Genomics 23:125-131(1994).
 CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
 CC NOCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
 CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
 CC INTO AN EXTENSIBLE 3D NETWORK.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U08210; AAA80155.1; -
 DR PIR; A55721; EAMS.
 DR MGD; MGI:95317; Eln.
 DR InterPro; IPR003979; tropoelastin.
 DR PRINTS; PR01500; TROP0ELASTIN.
 KW Structural protein; Repeat; Signal; Connective tissue.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 860 ELASTIN.
 FT DISULFID 850 855 BY SIMILARITY.
 SQ SEQUENCE 860 AA; 71955 MW; 0C0B55AAE1EDD7F1 CRC64;

Query Match 7.9%; Score 153.5; DB 1; Length 860;
 Best Local Similarity 24.4%; Pred. No. 0.069;
 Matches 115; Conservative 34; Mismatches 192; Indels 131; Gaps 17;
 QY 5 GALP-----PEINARMYAGPGSASLVAAQWDSVADLSFSAASAFQSVVWGLTVGSGW-- 58
 Db 402 GGIPGVGGPGICGGPGIVGGPGAVSPAATAKAAKAAK--YGAEGGVGITYGVGAGGPGG 459
 QY 59 --IGSAGLNVAAASPYVAVMSVTAQOELTAQVRAVAAAYETAYGLTVPPPVIAENRA 116
 Db 460 YGVGAGAGL--GGASPA--AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAG 491
 QY 117 ELMILITATNLGQNTPTALVNEAEVGEWMAODRAAMFGYAAATATATATATLLP-FEEAPEM 175
 Db 492 ALGGLVPGAVPGALPGAVPAVPGAGVPGAGTFAAAAAAATAKAAKAGLPGVGVGPGG 551
 QY 176 TSAGGL-----LEQAAVVEASDTAAANQ----- 199
 Db 552 VGVGGIPGGVGVGPGVGGVGGTIGAGPGGLGGAGSPAATAKAAKAAKAAKAAQYRAAG 611
 QY 200 LAMNVPAALQQLAQPTQGT-----TPSSKLGGLWKTIVSHRPSINMWSMANN 247
 Db 612 LGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPG 671


```

DR EMBL; M86355; AAA42271.1; JOINED.
DR EMBL; M86363; AAA42271.1; JOINED.
DR EMBL; M86364; AAA42271.1; JOINED.
DR EMBL; M86366; AAA42271.1; JOINED.
DR EMBL; M86371; AAA42271.1; JOINED.
DR EMBL; M86376; AAA42272.1; JOINED.
DR EMBL; M86373; AAA42272.1; JOINED.
DR EMBL; M86375; AAA42272.1; JOINED.
DR PIR; A36106; EART.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PRO1500; TROPOELASTIN.
KW Structural protein; Connective tissue; Repeat; Signal;
KW Alternative splicing.
FT NON_TER 1
FT SIGNAL <1 21
FT CHAIN 22 864
FT DISULFID 854 859
FT VARSPPLIC 263 307
FT
FT VARSPPLIC 308 308
FT
FT VARSPPLIC 809 823
FT
FT
SQ SEQUENCE 864 AA; 72786 MW; 456894BB09E79FD4 CRC64;

Query Match 7.7%; Score 150; DB 1; Length 864;
Best Local Similarity 25.9%; Pred. No. 0.11; Indels 108; Gaps 18;
Matches 102; Conservative 31; Mismatches 153;

QY 5 GALTPEINARMYAGPGSASLVAQAQWDSVASAFQSVVWGLTVGSW-- 58
Db 399 GGIPGVGGPGGIVGPGCAVSPAAAKAAKAAK--YCARGGVGIPTYGVGAGGPPG 456

QY 59 --IGSSAGLMVAASPYVAVMSVTAGQAEHTAAQURVAAAAYETAYGLTVPPVIAENRA 116
Db 457 YGVGAGAGL-----GGASQAAAA-----AAAKAAKYGAG----- 486

QY 117 ELMILIAITNLGQNTPAIVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEAPEMT 176
Db 487 -----GAGTIGGLVPG-----AVPGALPGAVPGALPGAVPGALP-GAVPGVP 527

QY 177 SAGLL-----EQAAVEEASDTAAANQLMNVVPAQLQAQPTQGTTPSS-KLGLWKTVS 232
Db 528 GTGGVPGAGTFAAAAAAATAKAA-----KAGQYGLGPGVGVGPGVGLPGGVG 579

QY 233 PHRSPISNNVSMANNHSMNTNSGVSM-TNTLSSMLKGF-APAAAAQAVQTAQNGVEAMS 290
Db 580 PG-----CVTGIGTGTGLVPGDLGGAGTPAAKSAKAAKAAKAYRAAA 624

QY 291 SLGSSLGSSGLGGV-----AANLGRAASV---GSLSVPOAWAANQAVTPAARALELT 341
Db 625 GLGAGVPCGLGVGAGVPGFAGAGGAGAGAGVPGFAGAGVPGSLAASKAAYGAAGL--- 681

QY 342 SLTSAERCPGCMGLGGLPVGMGARAG--GGLSGV 374
Db 682 -----CGPGGLGPGGLGPGFPGFPGGLGV 708

```

Search completed: November 21, 2003, 16:04:20
Job time : 6.61043 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 70.9578 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-52

Perfect score: 6061

Sequence: 1 MQHHHHHTDVSVGNLEIA.....SAATRRPCTGRDGRWACQ 1172

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19JUN03.*

```
1: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6061	100.0	1172	22 AAU01900	M. tuberculosis Tb
2	3866.5	63.8	2502	21 AAB00117	M. tuberculosis po
3	3810	62.9	741	22 AAE05665	Mycobacterium tube
4	2289	37.8	739	16 AAR77501	Malate synthase.
5	2289	37.8	739	22 AAG92334	C glutamicum prote
6	2275	37.5	739	22 AAB79539	Corynebacterium gl
7	2253	37.2	712	22 AAB79540	Corynebacterium gl
8	1619.5	26.7	421	22 AAU08233	Mycobacterium poly
9	1252	20.7	842	15 AAR63567	Feedback inhibito

10	1252	20.7	842	15	AAR63570	Feedback inhibito
11	1248.5	20.6	421	22	AAG64046	Corynebacterium th
12	1248	20.6	842	15	AAR63568	Feedback inhibito
13	1248	20.6	842	15	AAR63569	Feedback inhibito
14	1227.5	20.3	421	21	AAB29607	Modified Corynebac
15	1227.5	20.3	421	24	ABP97764	Amino acid sequenc
16	1224.5	20.2	420	24	ABP97763	Amino acid sequenc
17	1224.5	20.2	421	22	AAU71874	C. glutamicum meta
18	1224.5	20.2	421	22	AAG90027	C. glutamicum prote
19	1224.5	20.2	421	22	AAR79646	Corynebacterium gl
20	1223.5	20.2	421	15	AAR51469	AK alpha subunit T
21	1223.5	20.2	421	15	AAR51465	AK alpha subunit T
22	1223.5	20.2	421	16	AAR71416	C. glutamicum prot
23	1223.5	20.2	421	16	AAR63507	Corynebacterium gl
24	1223.5	20.2	421	19	AAW68147	Aspartokinase alph
25	1223.5	20.2	421	19	AAW69547	Brevibacterium lac
26	1223.5	20.2	421	19	AAW47401	B. lactofermentum
27	1219.5	20.1	421	14	AA540186	B. flavum AK. Bre
28	1219.5	20.1	421	15	AAR51466	AK wild type alpha
29	1219.5	20.1	421	15	AAR51470	AK wild type alpha
30	1219.5	20.1	421	18	AAW06586	B. lactofermentum
31	1219.5	20.1	421	19	AAW68152	Aspartokinase alph
32	1218.5	20.1	421	15	AAR63571	Feedback inhibito
33	1217.5	20.1	421	22	AAR04857	Corynebacterium gl
34	1217.5	20.1	421	23	ABR81982	C. glutamicum aspa
35	1215.5	20.1	421	15	AAR63566	Feedback inhibito
36	1211.5	20.0	421	22	AAE04867	Corynebacterium sp
37	1211.5	20.0	421	23	ABR81989	C. glutamicum aspa
38	1176	19.4	505	22	AAU55464	Propionibacterium
39	991.5	16.4	1190	22	ABG18357	Novel human diagno
40	787	13.0	409	21	ABG29516	Methylophilus meth
41	785	13.0	411	14	AA540846	Aspartokinase II.
42	773.5	12.8	254	23	ABP65555	Bifidobacterium lo
43	770	12.7	156	20	AAI39178	M. tuberculosis an
44	770	12.7	156	20	AAI39035	M. tuberculosis re
45	741.5	12.2	830	22	AAG25453	Novel human diagno

ALIGNMENTS

RESULT 1

AAU01900

ID AAU01900 standard; Protein; 1172 AA.

AC AAU01900;

DT 29-AUG-2001 (first entry)

DE M. tuberculosis TbF14 fusion protein.

KW TBFL4; antigen; vaccine; tuberculosis; AIDS; His tag; MtB81; Mo2;

KW acquired immunodeficiency disease.

OS Synthetic.

OS Mycobacterium tuberculosis.

XX Key

XX Binding-site

XX Location/Qualifiers

XX 3..8

XX /label= Histidine_tag

XX /note= "Nickel chelating region used for purifying

XX the fusion protein"

XX Region

XX 9..749

XX /label= MtB81_region

XX Misc-difference 49

XX /note= "Encoded by TC"

XX Region

XX 750..1172

XX /label= Mo2_region

XX WO200124820-A1.

XX 12-APR-2001.

XX

PF 10-OCT-2000; 2000WO-US28095.
 XX 07-OCT-1999; 99US-0158338.
 PR 07-OCT-1999; 99US-0158425.
 XX (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX WPI; 2001-290576/30.
 DR N-PSDB; AAS03794.
 XX
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 XX Claim 3; Fig 3; 16pp; English.
 XX
 XX The sequence represents Mycobacterium fusion protein antigen Tbf14
 CC consisting of a His tag for purification, antigen MtB81 and antigen Mo2.
 CC Compositions comprising at least 2 heterologous antigens, as a fusion
 CC protein, and vectors expressing the fusion proteins are used as vaccines
 CC to prophylactically immunise mammals (especially humans) against
 CC infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS.
 XX
 XX Sequence 1172 AA;
 SQ
 Query Match 100.0%; Score 6061; DB 22; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQHHHHHTDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNAL 60
 DB 1 MQHHHHHTDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNAL 60
 QY 61 LNARDELQAQIDKWHRRRVEIPIDMDAYRQFTEIGYLLPEPDDFTITTSVDABITTA 120
 DB 61 LNARDELQAQIDKWHRRRVEIPIDMDAYRQFTEIGYLLPEPDDFTITTSVDABITTA 120
 QY 121 GPOLVVPVNLNARFALNANRWGSLYDALYGTVDVPEPDAEKGPTYNKVRGDKVIAYAR 180
 DB 121 GPOLVVPVNLNARFALNANRWGSLYDALYGTVDVPEPDAEKGPTYNKVRGDKVIAYAR 180
 QY 181 KFLDDSVPLSSGSGFGDATGFTVQDGLVVALPKSTGLANPGQFAGYTGAESPTSVLLI 240
 DB 181 KFLDDSVPLSSGSGFGDATGFTVQDGLVVALPKSTGLANPGQFAGYTGAESPTSVLLI 240
 QY 241 NHGLHIEILLIDPESQVGTTRDAGVKDVILESATITIMDFEDSVAADVADKVLGYRNMLG 300
 DB 241 NHGLHIEILLIDPESQVGTTRDAGVKDVILESATITIMDFEDSVAADVADKVLGYRNMLG 300
 QY 301 LNKGDJAAAVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRSIMFVRNVGHLMTNDAIVDT 360
 DB 301 LNKGDJAAAVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRSIMFVRNVGHLMTNDAIVDT 360
 QY 361 DGSEVFEIGIMDALFTGLIAIHGLKASDVNGPLINSRSGSIYIVKPRMHPAEVATFCELF 420
 DB 361 DGSEVFEIGIMDALFTGLIAIHGLKASDVNGPLINSRSGSIYIVKPRMHPAEVATFCELF 420
 QY 421 SRVEDVLGHPONTMKIGIMDEERTTNLKCACIAAADRVVFTINTGFLDRTGDEIHTSWE 480
 DB 421 SRVEDVLGHPONTMKIGIMDEERTTNLKCACIAAADRVVFTINTGFLDRTGDEIHTSWE 480
 QY 481 AGPMVRKGTWKSQFWILAYEDHNVDAGLAAGSGRAQVKGWMTWMTLMADVMVETKIAQP 540
 DB 481 AGPMVRKGTWKSQFWILAYEDHNVDAGLAAGSGRAQVKGWMTWMTLMADVMVETKIAQP 540
 QY 541 RAGASTAWPSPPTAATLHALHVOVDVAVVOOGLAGKRRTATIEQLLTIPLAKELAWAPDE 600
 DB 541 RAGASTAWPSPPTAATLHALHVOVDVAVVOOGLAGKRRTATIEQLLTIPLAKELAWAPDE 600

QY 601 IREVDNNCSILGYVVRWVDQGVGSKVPDIIHDVALMEDRATLRISSQLLANLWLRHGV 660
 DB 601 IREVDNNCSILGYVVRWVDQGVGSKVPDIIHDVALMEDRATLRISSQLLANLWLRHGV 660
 QY 661 TSADVRLASLERMAPLVDRQNAVGVAYRPNAPNDDSIAPLAQELILSGAQQNGYTEPI 720
 DB 661 TSADVRLASLERMAPLVDRQNAVGVAYRPNAPNDDSIAPLAQELILSGAQQNGYTEPI 720
 QY 721 LHRREFKARAAEKAPSDRAGDDAARVQKYGSSVDAERIRRVAAERIVATKKGNDV 780
 DB 721 LHRREFKARAAEKAPSDRAGDDAARVQKYGSSVDAERIRRVAAERIVATKKGNDV 780
 QY 781 VVVVSAMGDTTDDLLDAQVCPAPPRELDMLLTAGERISNALVAMATESLGAHARSFT 840
 DB 781 VVVVSAMGDTTDDLLDAQVCPAPPRELDMLLTAGERISNALVAMATESLGAHARSFT 840
 QY 841 GSAQGVITTTGTHGNAKIIDVTPGRLQTALEEGRWLVAGFQGVSDTKDVTTLGRGSDT 900
 DB 841 GSAQGVITTTGTHGNAKIIDVTPGRLQTALEEGRWLVAGFQGVSDTKDVTTLGRGSDT 900
 QY 901 TAVAMAALGADYCEIYTDVVGIFSDADPRIVRNARKLDTVTPEEMLEMAACGAKVLMRC 960
 DB 901 TAVAMAALGADYCEIYTDVVGIFSDADPRIVRNARKLDTVTPEEMLEMAACGAKVLMRC 960
 QY 961 VEYARRHNI PVHVRSSYSYSDRPGTVVVGSIKDVPMEDPILTVGAHDSRSEAKVTIVGLPDIP 1020
 DB 961 VEYARRHNI PVHVRSSYSYSDRPGTVVVGSIKDVPMEDPILTVGAHDSRSEAKVTIVGLPDIP 1020
 QY 1021 GYAAKVPFVARRRRQHRHGAERLOGRQDRHHLHLIPQTSGLPPPPKWNTRSETRSAS 1080
 DB 1021 GYAAKVPFVARRRRQHRHGAERLOGRQDRHHLHLIPQTSGLPPPPKWNTRSETRSAS 1080
 QY 1081 TOLLYDDHIKGVSLIGAGVSRHPGVTATFCEALAAVGVNIELISTSEDORSRCCAATPNW 1140
 DB 1081 TOLLYDDHIKGVSLIGAGVSRHPGVTATFCEALAAVGVNIELISTSEDORSRCCAATPNW 1140
 QY 1141 TRPMSRCMKSSGSAATRRPRCTRGRDGRWACQ 1172
 DB 1141 TRPMSRCMKSSGSAATRRPRCTRGRDGRWACQ 1172

RESULT 2
 AAB00117
 ID AAB00117 standard; Protein; 2502 AA.
 XX AAB00117;
 AC AAB00117;
 XX
 XX 08-FEB-2001 (first entry)
 DT
 XX M. tuberculosis polypeptide sequence comprising Mtb-81 antigen.
 DE
 XX Mycobacterium tuberculosis; antigen; Mtb-91; Mtb-67.2; APC;
 KW antigen presenting cell; serodiagnosis; detection;
 KW human immunodeficiency virus; HIV;
 KW acquired immune deficiency syndrome; AIDS.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 76 /note= "Encoded by TAG stop codon"
 FT
 FT Misc-difference 87 /note= "Encoded by TGA stop codon"
 FT
 FT Misc-difference 88 /note= "Encoded by TAG stop codon"
 FT
 FT Misc-difference 124 /note= "Encoded by TAA stop codon"
 FT
 FT Misc-difference 140 /note= "Encoded by TAA stop codon"
 FT
 FT Misc-difference 174 /note= "Encoded by TAA stop codon"
 FT
 FT Misc-difference 185 /note= "Encoded by TAA stop codon"
 FT

FT Misc-difference 216 /note= "Encoded by TAA stop codon"
FT Misc-difference 217 /note= "Encoded by TGA stop codon"
FT Misc-difference 227 /note= "Encoded by TAG stop codon"
FT Misc-difference 248 /note= "Encoded by TAA stop codon"
FT Misc-difference 259 /note= "Encoded by TAA stop codon"
FT Misc-difference 267 /note= "Encoded by TAA stop codon"
FT Misc-difference 270 /note= "Encoded by TAA stop codon"
FT Misc-difference 305 /note= "Encoded by TAA stop codon"
FT Misc-difference 340 /note= "Encoded by TAA stop codon"
FT Misc-difference 346 /note= "Encoded by TAA stop codon"
FT Misc-difference 360 /note= "Encoded by TAA stop codon"
FT Misc-difference 387 /note= "Encoded by TAG stop codon"
FT Misc-difference 423 /note= "Encoded by TGA stop codon"
FT Misc-difference 436 /note= "Encoded by TGA stop codon"
FT Misc-difference 444 /note= "Encoded by TAA stop codon"
FT Misc-difference 475 /note= "Encoded by TGA stop codon"
FT Misc-difference 492 /note= "Encoded by TAG stop codon"
FT Misc-difference 507 /note= "Encoded by TAA stop codon"
FT Misc-difference 536 /note= "Encoded by TAA stop codon"
FT Misc-difference 550 /note= "Encoded by TAA stop codon"
FT Misc-difference 555 /note= "Encoded by TAG stop codon"
FT Misc-difference 563 /note= "Encoded by TAG stop codon"
FT Misc-difference 572 /note= "Encoded by TAA stop codon"
FT Misc-difference 623 /note= "Encoded by TGA stop codon"
FT Misc-difference 647 /note= "Encoded by TAA stop codon"
FT Misc-difference 750 /note= "Encoded by TAA stop codon"
FT Misc-difference 758 /note= "Encoded by TGA stop codon"
FT Misc-difference 784 /note= "Encoded by TGA stop codon"
FT Misc-difference 816 /note= "Encoded by TGA stop codon"
FT Misc-difference 854 /note= "Encoded by TAA stop codon"
FT Misc-difference 861 /note= "Encoded by TGA stop codon"
FT Misc-difference 917 /note= "Encoded by TAA stop codon"
FT Misc-difference 1035 /note= "Encoded by TAA stop codon"
FT Misc-difference 1082 /note= "Encoded by TAA stop codon"
FT Misc-difference 1160 /note= "Encoded by TGA stop codon"
FT Misc-difference 1161 /note= "Encoded by TAA stop codon"
FT Misc-difference 1164 /note= "Encoded by TGA stop codon"

FT Misc-difference 1187 /note= "Encoded by TAA stop codon"
FT Misc-difference 1188 /note= "Encoded by TAA stop codon"
FT Misc-difference 1211 /note= "Encoded by TGA stop codon"
FT Misc-difference 1253 /note= "Encoded by TGA stop codon"
FT Misc-difference 1344 /note= "Encoded by TGA stop codon"
FT Misc-difference 1348 /note= "Encoded by TAA stop codon"
FT Misc-difference 1386 /note= "Encoded by TAA stop codon"
FT Misc-difference 1527 /note= "Encoded by TAA stop codon"
FT Misc-difference 1579 /note= "Encoded by TAA stop codon"
FT Misc-difference 1582 /note= "Encoded by TGA stop codon"
FT Misc-difference 1616 /note= "Encoded by TGA stop codon"
FT Misc-difference 1628 /note= "Encoded by TGA stop codon"
FT Misc-difference 1643 /note= "Encoded by TGA stop codon"
FT Misc-difference 1673 /note= "Encoded by TAA stop codon"
FT Misc-difference 1683 /note= "Encoded by TAA stop codon"
FT Misc-difference 1685 /note= "Encoded by TAA stop codon"
FT Protein 1691..2438 /label= Mtb-81 polypeptide
FT Misc-difference 2439 /note= "Encoded by TAG stop codon"
FT Misc-difference 2440 /note= "Encoded by TGA stop codon"
FT Misc-difference 2464 /note= "Encoded by TAA stop codon"
FT Misc-difference 2471 /note= "Encoded by TGA stop codon"
FT Misc-difference 2478 /note= "Encoded by TGA stop codon"
FT Misc-difference 2489 /note= "Encoded by TAA stop codon"
XX /note= "Encoded by TAA stop codon"
PN WO200055194-A2.
XX 21-SEP-2000.
XX 17-MAR-2000; 2000WO-US07196.
XX 18-MAR-1999; 99US-0272975.
XX (CORI-) CORIXA CORP.
XX Hendrickson RC, Lodes WJ, Houghton RL;
XX WPI; 2000-638180/61.
XX N-PSDB; AAA53971.
XX Novel Mycobacterium tuberculosis polypeptide comprising an immunogenic
PT portion of M. tuberculosis antigens Mtb-81 and Mtb-67.2, useful for
PT diagnosis, treatment and monitoring therapy of tuberculosis
XX Claim 1; Fig 1a-f; 91pp; English.
XX Polypeptides comprising an immunogenic portion of Mycobacterium
CC tuberculosis antigens Mtb-81 or Mtb-67.2 or variants that differ
CC by substitutions, additions, insertions and/or deletions but which
CC still react with antigen specific antisera or T-cells are described.

CC Also described are polynucleotides encoding the polypeptides. The
CC polypeptides, expression vectors expressing the polypeptides or
CC comprising an antisense polynucleotide, or an antigen presenting
CC cell comprising a sequence encoding the polypeptides are useful for
CC determining the presence or absence of M tuberculosis in whole blood,
CC serum, sputum, plasma, saliva, cerebrospinal fluid or urine in a
CC patient infected with human immunodeficiency virus (HIV).
XX
XX
XX Sequence 2502 AA;
Query Match 63.8%; Score 3866.5; DB 21; Length 2502;
Best Local Similarity 88.3%; Pred. No. 3.8e-286;
Matches 768; Conservative 8; Mismatches 27; Indels 67; Gaps 6;
QY 1 MQHHHHHTDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNAL 60
DB 1691 MQHHHHHTDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNAL 1750
QY 61 LNARDELQAIQDKWHRVRIEIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTA 120
DB 1751 LNARDELQAIQDKWHRVRIEIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTA 1810
QY 121 GPOLVVPVNLNARFALNAARNGSLYDALYGTDPETDGAEGKGTYNKVRGDKVIAYAR 180
DB 1811 GPOLVVPVNLNARFALNAARNGSLYDALYGTDPETDGAEGKGTYNKVRGDKVIAYAR 1870
QY 181 KFLDDSVPLSSGSGFGDATGFTVDGQLVVALPKSTGLANPGQFAGYTGAESPVSLLI 240
DB 1871 KFLDDSVPLSSGSGFGDATGFTVDGQLVVALPKSTGLANPGQFAGYTGAESPVSLLI 1930
QY 241 NHGLHIEILLDPBSQVGTTRAGVKDVIIESAITTIMDPEDSVAAVDAADKVLGYRNWLG 300
DB 1931 NHGLHIEILLDPBSQVGTTRAGVKDVIIESAITTIMDPEDSVAAVDAADKVLGYRNWLG 1990
QY 301 LNKGDIAAAVDKDGTAFLVNLNDRNYTAPGGQFTLPORSIMFVRNVLMTNDAIVDT 360
DB 1991 LNKGDIAAAVDKDGTAFLVNLNDRNYTAPGGQFTLPORSIMFVRNVLMTNDAIVDT 2050
QY 361 DGSSEFEGINDALFTGLIAHGKASDVNGPLNSRTGSIYIVKPMHGPFAVFTCELF 420
DB 2051 DGSSEFEGINDALFTGLIAHGKASDVNGPLNSRTGSIYIVKPMHGPFAVFTCELF 2110
QY 421 SRVEDVLGLPONTMKIGIMDEERTTNLKCACIAAADRWFINTGFLDRTGDEIHTSME 480
DB 2111 SRVEDVLGLPONTMKIGIMDEERTTNLKCACIAAADRWFINTGFLDRTGDEIHTSME 2170
QY 481 AGPMVRKGTMSQFWILAYEDHNVDAAGLAFSGRAQVKGKMTMELNADVETKIAQP 540
DB 2171 AGPMVRKGTMSQFWILAYEDHNVDAAGLAFSGRAQVKGKMTMELNADVETKIAQP 2230
QY 541 RAGASTAWVPSPTAATILHLYHQVDVAAVQOQGLAGKRRATIEQLLTIPLAKELAWAPDE 600
DB 2231 RAGASTAWVPSPTAATILHLYHQVDVAAVQOQGLAGKRRATIEQLLTIPLAKELAWAPDE 2290
QY 601 IREVDNCCOSILGYVVRVWDQVGCSPDIDHDVALMEDRATLRISQQLLANLWLBHGYI 660
DB 2291 IREVDNCCOSILGYVVRVWDQVGCSPDIDHDVALMEDRATLRISQQLLANLWLBHGYI 2350
QY 661 TSADVRLASLERMAPLVDQVAGVAVRPMAPNFDDSIPLAQLILSGAQPNGYTEPI 720
DB 2351 TSADVRLASLERMAPLVDQVAGVAVRPMAPNFDDSIPLAQLILSGAQPNGYTEPI 2410
QY 721 LHRRRREFKARAEKAPSDRAGDAAARVOKYGGSSVDAERIRRVARIVATKOGNDV 780
DB 2411 LHRRRREFKARAEKAPSDRAGDAAARVOKYGGSSVDAERIRRVARIVATKOGNDV 2440
QY 781 VVVVSAMGDTTDDLLDLAQVCPAPPPELMDLLTAGERISNALVAVAIESLGAHARSFT 840
DB 2441 LQIS-----ITLAAARAPPPP-----LRSG-----CXQSPK 2468
QY 841 GSAQGVITTTGTHGNAKIIDVTPGRLQTALE 870
DB 2469 GSXVGC-----CHRXA-ITSITPWLXTGLE 2493

RESULT 3
AAE05665
ID AAE05665 standard; peptide; 741 AA.
XX
AC AAE05665;
XX
DT 24-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis 88 kDa secreted protein.
XX
KW Mycobacterial disease; infection; 88-kDa protein; tuberculosis; TB;
KW early antigen; human immunodeficiency virus; HIV; secreted protein.
XX
OS Mycobacterium tuberculosis.
XX
PN US6245331-B1.
XX
PD 12-JUN-2001.
XX
PF 31-DEC-1997; 97US-0001984.
XX
PR 02-JAN-1997; 97US-0034003.
XX
PA (UJNY-) UNIV NEW YORK MEDICAL CENT.
XX (COLS) UNIV COLORADO STATE.
XX
PI Laal S, Zolla-Pazner S, Belisile JT;
XX
XX WPI; 2001-424324/45.
XX
XX Detecting a mycobacterial disease (tuberculosis) in individuals
PT comprise assaying a biological sample for the presence of
PT anti-Mycobacterium tuberculosis antibodies or M. tuberculosis
PT antigen-antibody complex -
XX
PS Claim 14; Column 121; 96pp; English.
XX
XX The present invention relates to a method for early detection of active
XX mycobacterial disease or infection comprising assaying a biological fluid
XX sample for the presence of early antibodies specific for an 88-kDa
XX Mycobacterium tuberculosis (Mt) protein or immune complexes consisting of
XX an 88-kDa M. tuberculosis protein antigen complexed with an antibody
XX specific for the antigen. The method is useful for the early and rapid
XX detection of mycobacterial disease, particularly tuberculosis, in
XX individuals at heightened risk of developing tuberculosis. This
XX individuals include human immunodeficiency virus (HIV)-infected subjects
XX or other immunocompromised individuals. The method is a rapid and
XX inexpensive screening procedure for detecting mycobacterial disease.
XX The present sequence is a Mt 88 kDa secreted protein early antigen,
XX used in the exemplification of the invention.
XX
XX Sequence 741 AA;

Query Match 62.9%; Score 3810; DB 22; Length 741;
Best Local Similarity 100.0%; Pred. No. 1.3e-282;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 TDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNALNARDELQ 68
DB 2 TDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNALNARDELQ 61
QY 69 AQIDKWHRRVRIEIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTAGPOLVVPV 128
DB 62 AQIDKWHRRVRIEIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTAGPOLVVPV 121
QY 129 LNARFALNAARNGSLYDALYGTDPETDGAEGKGTYNKVRGDKVIAYARFLDDSV 188
DB 122 LNARFALNAARNGSLYDALYGTDPETDGAEGKGTYNKVRGDKVIAYARFLDDSV 181
QY 189 LSSGSFGDATGFTVDGQLVVALPKSTGLANPGQFAGYTGAESPVSLLINHLHIEI 248

Db 182 LSSGSGDAGTGTQDQGVVALPKSTGLANPGQAGYTGAEPSVLLINHLGHIEI 241
 Qy 249 LIDPSSQVGTTRAGVKDVILESATTTIMDFEDSVAADADKVLGYRNWGLNKGDLAA 308
 Db 242 LIDPSSQVGTTRAGVKDVILESATTTIMDFEDSVAADADKVLGYRNWGLNKGDLAA 301
 Qy 309 AVDKDGTAFRLVLRNDRNTAPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDGSEVFEG 368
 Db 302 AVDKDGTAFRLVLRNDRNTAPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDGSEVFEG 361
 Qy 369 IMDALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKHGPAPAEVAFCELFSEVEDVLG 429
 Db 362 IMDALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKHGPAPAEVAFCELFSEVEDVLG 421
 Qy 429 LPQNTMKIGIMDEERTITVNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGPMVRKG 488
 Db 422 LPQNTMKIGIMDEERTITVNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGPMVRKG 481
 Qy 489 TKKSQPWILAYEDHNVDAGLAAGFSGRAQVKGKMWMTLMADWVETKIAOPRAGASTAW 548
 Db 482 TKKSQPWILAYEDHNVDAGLAAGFSGRAQVKGKMWMTLMADWVETKIAOPRAGASTAW 541
 Qy 549 VPSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDEIREEVN 608
 Db 542 VPSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDEIREEVN 601
 Qy 609 CQSILGYVVRVVDQVGCSPKVPDIHDVALMEDRATLRISQLLANWLHGHVITSADYRAS 668
 Db 602 CQSILGYVVRVVDQVGCSPKVPDIHDVALMEDRATLRISQLLANWLHGHVITSADYRAS 661
 Qy 669 LERMAPLVDRQNVAGDVAPRMAPNFDSDIAFLAQELILSGAQOPNGYTEPILHRRREF 728
 Db 662 LERMAPLVDRQNVAGDVAPRMAPNFDSDIAFLAQELILSGAQOPNGYTEPILHRRREF 721
 Qy 729 KARAAEKAPSDRAGDDAAR 748
 Db 722 KARAAEKAPSDRAGDDAAR 741

RESULT 4

AAR77501
 ID AAR77501 standard; Protein; 739 AA.

XX AC AAR77501;

XX DT 04-JUN-1996 (first entry)

XX DE Malate synthase.

XX KW aceB gene; malate synthase; Coryneform bacterium; regulatory region;
 XX KW protein synthesis.

OS Coryneform glutamicum.

XX DE4440118-C1.

XX PD 09-NOV-1995.

XX PF 11-NOV-1994; 94DE-4440118.

XX PR 11-NOV-1994; 94DE-4440118.

XX PA (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.

XX PI Eikmanns B, Reinscheid D, Sahn H;

XX DR WPI; 1995-374802/49.

XX DR N-PSDB; AAT05501.

XX DNA contg. upstream regulatory region from malate synthase gene of
 PT coryneform bacteria - used for regulated synthesis of protein in
 PT coryneform(s), regardless of nutrient medium compen..

PS Example 3; Page 7-8; 12pp; German.

XX CC AAR77501 is encoded by an approx. 3 kb BfrI-PvuI fragment contg. the
 CC aceB gene from Coryneform glutamicum (see AAT05501). Nucleotides 1
 CC to 574 can be isolated and used to regulate the expression of a
 CC protein-encoding structural gene placed downstream of it, after
 CC incorporation into a vector and transfer into a Coryneform bacterium.
 CC Induction of protein expression occurs even when the culture medium
 CC contains carbon sources other than acetate.

SQ Sequence 739 AA;

Query Match 37.8%; Score 2289; DB 16; Length 739;
 Best Local Similarity 61.3%; Pred. No. 3.6e-166;
 Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;

Qy 9 TDRVSVGNRIARVLVDFVNNALPCTDIDPDSFWAGVDKVVADLTPOQNALINARDELQ 69
 Db 20 TERVDAGNQVAKVLYDFVTEAVLPRVGVDAEKFWGFAAARDLTPRRELLARDELQ 79
 Qy 69 AQIDKWHRRRVIEPIDMDAYROFLTEIGVLLPEPDDFTITSGVDAEITTTAGPQLVVPV 128
 Db 80 MLIDDDYHRNN-SGTIDQEAYEDFLKEIGVLYVEPEAAEIRTQNVDTIEISSTAGPQLVVP 138
 Qy 129 LNAARFALNANARWGLYDALYDGTVDVTPETDGAEGKFTYKVRGDKVIAYARKFLDDSV 188
 Db 139 LNAARFALNANARWGLYDALYGTNAIPETDGAEGKFTYKVRGDKVIAYARKFLDDSV 198
 Qy 189 LSSGSGDAGTGTQDQGVVALPKSTGLANPGQAGYTGAEPSVLLINHLGHIEI 248
 Db 199 LDGASHADVEKNITDGLAAHIGDSVYRLKRESYRGFTGNFLDPEAILLETNGHIL 258
 Qy 249 LIDPSSQVGTTRAGVKDVILESATTTIMDFEDSVAADADKVLGYRNWGLNKGDLAA 308
 Db 259 QIDPVHPIGKADTKGLKDIVLESATTTIMDFEDSVAADADKVLGYRNWGLNKGDLAA 318
 Qy 309 AVDKDGTAFRLVLRNDRNTAPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDGSEVFEG 368
 Db 319 EMSKNGRIPTRELKNDRVYIGRNGTELVLHGRSLLFVRNVGHLMTNDAIVDTDGSEVFEG 377
 Qy 369 IMDALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKHGPAPAEVAFCELFSEVEDVLG 428
 Db 378 IMDALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKHGPAPAEVAFCELFSEVEDVLG 434
 Qy 429 LPQNTMKIGIMDEERTITVNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGPMVRKG 488
 Db 435 LPRHTLKVGVMDEERTITVNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGPMVRKG 494
 Qy 489 TKKSQPWILAYEDHNVDAGLAAGFSGRAQVKGKMWMTLMADWVETKIAOPRAGASTAW 548
 Db 495 DMQTAPWKQAYENNVNVDAGIQRLPKAQIGKGMWMTLMADWVETKIAOPRAGASTAW 554
 Qy 549 VPSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDEIREEVN 607
 Db 555 VPSPTGATLHATHYHLVDVFKVQDELRAAGRDRSLRNILTIPTAPNTNWESEKKEMDN 614
 Qy 608 NCOSILGYVVRVVDQVGCSPKVPDIHDVALMEDRATLRISQLLANWLHGHVITSADYRAS 667
 Db 615 NCOSILGYVVRVVEHGVGCSPKVPDIHDIDIMEADRATLRISQLLANWLHGHVITSADYRAS 674
 Qy 668 SLERMAPLVDRQNVAGDVAPRMAPNFDSDIAFLAQELILSGAQOPNGYTEPILHRRREF 727
 Db 675 SLERMAPLVDRQNVAGDVAPRMAPNFDSDIAFLAQELILSGAQOPNGYTEPILHRRREF 734
 Qy 728 FKAR 731
 Db 735 FKAR 738

RESULT 5
 AAG93234
 ID AAG93234 standard; Protein; 739 AA.
 XX

PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1032180.
PR 08-JUL-1999; 99DE-1032227.
PR 08-JUL-1999; 99DE-1032230.
PR 08-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX (BADI) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
PI WPI; 2001-061975/07.
DR N-PSDB; AAF71657.
DR
XX
XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
XX Claim 20; Page 998-1000; 1246pp; English.
XX
XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carboxylate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
XX Sequence 712 AA;
XX
XX Query Match 37.2%; Score 2253; DB 22; Length 712;
XX Best Local Similarity 61.2%; Pred. No. 1.9e-163;
XX Matches 438; Conservative 89; Mismatches 183; Indels 6; Gaps 4;
XX
XX 17 LR1ARVLYDFVNEALPGTIDPDSWAGVKKVADLTQNOALLNARDELQADKWRH 76
XX 1 MQVAKVLYDFVTEAVLPVGVDAKEFWGFAALRDLTPRNEKLLARRDELQMLDDYHR 60
XX
XX 77 RRVIEFDMDAYRQFTEIGYLLPEDDFTITTSYGDABEITTAGPQLVVPVYNARFALN 136
XX 61 NN-SGRTDQAEYDFLKEIGYLVVEEPAEAINTQNVDTISSTAGPQLVVPVYNARFALN 119
XX
XX 137 ANARWGSGLYDALYGTGDPETDGAEGKTYNKGDKVIAYARKFLDSDVPLSGSGCD 196
XX 120 ANARWGSGLYDALYGINAIPETDGAEGKTYNKGDKVIAYARKFLDSDVPLSGSGCD 179

QY 197 ATGFTVQDQGLVVALPKSTGLANPCQFAGYTGAAARSPTSVLINHLHIEILIDRESQV 256
Db 180 VEKYNITDGLAAHAGSVYRLKXNRSYRGFTCNFLDPEAILLETNGLHIEIQDPVPHI 239
QY 257 GTTDRAGVKDVILESATITIMDPEDSVAADAAKVLGYRNWGLNKLKGLDIAAAYDKDGA 316
Db 240 GKADKTGLKDIVLESATITIMDPEDSVAADAAEDKTLGYSNWFGLANTGELKEEWSXNGRI 299
QY 317 FLRVNDRNYTAPGGQFTLPGRSLMFVNVNCHLMTNDAIVDTDGSVEPEGIMDALFTG 376
Db 300 FTRELNKDRVYIGRNGTELVHGRSLLFVNVNCHLMTNDAIVDTDGSVEPEGIMDALFTG 358
QY 377 LIAIHGKASDVNGPLNSRTGSIYIVKPKWGHGPAEVAFTCELFPSVEDVLGPQNTMKI 436
Db 359 VCAIPGIAPQN---KWRNSRKGSIYIVKPKQHFEEVAFNLFGRVEDLDDLPRTLKV 415
QY 437 GIMDEERTTNLAKACIKAAAADRVFINTGFLDRTGDEIHTSMAGPMVRKGTMSQPMWI 496
Db 416 GVMDEERTSVNLDAIMEVADRLAFINTGFLDRTGDEIHTSMAGPMVRKGTMSQPMWI 475
QY 497 LAYEDHNVDAAGAFSGRAOVGKGMWMTETELXADMVETKIAOPRAGASTAWPSPSTAAT 556
Db 476 QAYENNVDAGIQRGELPKQAQIGKGMWMTETELXADMVETKIAOPRAGASTAWPSPSTAAT 535
QY 557 LHALHYHQVDVAAVQOGL-AGKERATTEQLLTIPLAKELAWAPDEIREEVDNNCQSILGY 615
Db 536 LHAETHYHLVDVFKVQDELRAAGRRDSLRNLTITSTAPNTNWSEKKEEDNNCQSILGY 595
QY 616 VVRVWDGVCCKVPDIHDVAMDEDRATLRISSQLLANWLHGVITTSADVRASLERMAPL 675
Db 596 VVRVWEGVCCKVPDIHDVAMDEDRATLRISSQLLANWLHGVITTSADVRASLERMAPL 655
QY 676 VDRQAGDVAYRPMAPNFDSDIAFLAAQELILSGAOPNGYTPILHRRRERFEKAR 731
Db 556 VDKQVAGDEAYRDMAPKYDASLAFQAQAKDLIFGTSKSPSGYTEPILHARRERFEKAK 711

XX
XX RESULT 8
XX AAU08233
XX ID AAU08233 standard; Protein; 421 AA.
XX AC AAU08233;
XX DT 17-DEC-2001 (first entry)
XX DE Mycobacterium polypeptide for MO-2.
XX KW Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
XX immunostimulant; MO-2.
XX OS Mycobacterium sp.
XX FH Key Location/Qualifiers
XX FT Misc-difference 41 /note= "Encoded by T"
XX FT Misc-difference 305 /note= "Encoded by GAC"
XX FT
XX PN WO200162893-A2.
XX PD 30-AUG-2001.
XX PF 26-FEB-2001; 2001WO-US05992.
XX PR 25-FEB-2000; 2000US-0185037.
XX PR 08-AUG-2000; 2000US-0223828.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Skeiky Y, Ovendale P, Jen S, Lodes M;
XX WPI; 2001-536638/59.
XX DR


```
QY 863 GLQTALIEGRVVLVAGFGVQSODTKDVTTLGRGSDTTAVAAALGADVCEIYTDVGG 922
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 540 GRVREALDEGKICIVAGFGVNKETRDVTTLGRGSDTTAVAAALGADVCEIYSDVGG 599
QY 923 IFSADPRIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSYSDRPG 982
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 600 VYTADPRIVPNAQKLEKLSFEEMLEMAAGSKILVLRSEYARAFNVPLRVRSYSDNPG 659
QY 983 TVVVGSIKDVPMEDPILTGVAHDRSEAKVTIVGLPDI PGYAAKVFRAVA-----R 1032
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 660 TLIAGSMEDIPVEAVLTGVAIDKSAKVTVLGISDKPGEAAKVFRALADAINDMVLQ 719
QY 1033 RRRQHRHGAERLQGRQD-RHHLHLP--QTSGPPPWKNWTRSETRSASTOLLYDDHI 1089
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 720 NVSSVEDGTTDITFTCPRSDGRAMEILKQLVQG-----NW-----TNVLYDDQV 765
QY 1090 GKVSILGAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 766 GKVSILGAGMRSHPGVTATFCEALAAVGVNIELISTSE 803

RESULT 11
AAG64046
ID AAG64046 standard; Protein; 421 AA.
XX
AC AAG64046;
XX
DT 11-SEP-2001 (first entry)
XX
DE Corynebacterium thermoaminogenes lysin biosynthetic enzyme lysC.
XX
KW Heat-resistant; lysin biosynthesis; enzyme; coryneform;
XX
KW aspartate-semialdehyde dehydrogenase; lysC.
XX
OS Corynebacterium thermoaminogenes.
```

```
Db 60 -----VPPAREMDMLLTAG-----ERISNALV 81
QY 375 TGLIAIHGLKASDVNGPLINSRGTGIYVKPKHGPAAVNAFTCELSRVEDVL----- 427
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 82 AMAIESLGAAQAQFTG-----SQAG---VLTERHGVARIVDVTP--GRVREALDEGKICI 132
QY 428 -----GLPQNTMKIGINDE--ERRTTVNLKACIKA-----AADRVVFINTGP 467
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 133 VAGPQGVNKETRDVTTLGRGSDTTAVAAALGADVCEIYSDVGVYADPRIVPNAQK 192
QY 468 LDRTGDIHHTSMEA-GPMVRKGMKQOPWILAY-----EDHNVDAAG-LAAGFSGRAQV 518
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 193 LEKLSFEEMLEMAAGSKIL--VLRSEYARAFNVPLRVRSYSDNPGTLIAGSMEDIPV 250
QY 519 GKGMWNTLMADMVETKIAQPRAGASTAVVPSPTAATLHALHYHQVDVAVQOGLAKGR 578
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 251 EEA--LTVGVAIDKSAKVTV--LGISDK--PGEAAKVFRALADAINDMVLQNVSSVE 304
QY 579 RATIEQLLTIP-----LAKELAWAPDEIRREYDNNCQSILGYVVRWVQ----- 622
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 305 DGTITDITFTCPRSDGRAMEILKKL-----OVQGNWTVL-----YDDQVGVKSLV 350
QY 623 GVGSKVPDIHDVALMEDRATLRISOLLANLRLRHGVITISADVRASLERMAPLVDRQAG 682
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 351 GAGMKSHPGV-TAEFMALRDVNVNIEL-----ISTSEIRIS-----VLIREDDL 394
QY 683 DVAYRMAPNPDSDSIAPLAAQELILSGAQPNGYTPILHRRRRREFKARAAKPAFSDRA 742
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 395 DAAARALHEQF-----QLGGEDEAVVYA-----GT 419
QY 743 GDDAARYQYKGGSVADAERIRRAERI VATKQGNNDVVVWSAMGDTTDLDDLQAOVC 802
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 420 GRVALVQYKGGSLAESAEIRNVARI VATKAGNNVVVVCAMGDTTDLLELAAYN 479
QY 803 PAPPRELDMLLTAGERISNALVAMAEISLGAHARFTGSQAGVITGTGHNKAKIIDVTP 862
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 480 VPPAREMDMLLTAGERISNALVAMAEISLGAHARFTGSQAGVITGTGHNKAKIIDVTP 539
QY 863 GLQTALIEGRVVLVAGFGVQSODTKDVTTLGRGSDTTAVAAALGADVCEIYTDVGG 922
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 540 GRVREALDEGKICIVAGFGVNKETRDVTTLGRGSDTTAVAAALGADVCEIYSDVGG 599
QY 923 IFSADPRIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSYSDRPG 982
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 600 VYTADPRIVPNAQKLEKLSFEEMLEMAAGSKILVLRSEYARAFNVPLRVRSYSDNPG 659
QY 983 TVVVGSIKDVPMEDPILTGVAHDRSEAKVTIVGLPDI PGYAAKVFRAVA-----R 1032
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 660 TLIAGSMEDIPVEAVLTGVAIDKSAKVTVLGISDKPGEAAKVFRALADAINDMVLQ 719
QY 1033 RRRQHRHGAERLQGRQD-RHHLHLP--QTSGPPPWKNWTRSETRSASTOLLYDDHI 1089
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 720 NVSSVEDGTTDITFTCPRSDGRAMEILKQLVQG-----NW-----TNVLYDDQV 765
QY 1090 GKVSILGAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 766 GKVSILGAGMRSHPGVTATFCEALAAVGVNIELISTSE 803

RESULT 11
AAG64046
ID AAG64046 standard; Protein; 421 AA.
XX
AC AAG64046;
XX
DT 11-SEP-2001 (first entry)
XX
DE Corynebacterium thermoaminogenes lysin biosynthetic enzyme lysC.
XX
KW Heat-resistant; lysin biosynthesis; enzyme; coryneform;
XX
KW aspartate-semialdehyde dehydrogenase; lysC.
XX
OS Corynebacterium thermoaminogenes.
```



```

XX JF2001120270-A.
XX PN
XX XX
XX PD
XX PF
XX PF
XX PR
XX PR
XX (AJIN ) AJINOMOTO KK.
XX PA
XX PA
XX WPI; 2001-364760/38.
XX N-PSDB; AAH45374.
XX DR
XX DR
XX A heat-resistant lysin biosynthetic system enzyme gene of a high
XX PT temperature-resistant coryneform microbe
XX PS Claim 6; Page 21-22; 27pp; Japanese.
XX PS
XX CC The invention relates to a gene from a high temperature-resistant
XX CC coryneform microbe that encodes a heat-resistant lysin biosynthetic
XX CC enzyme. The enzyme has aspartate-semialdehyde dehydrogenase activity
XX CC and can be used for growing amino acid-producing microbes. The
XX CC present amino acid sequence corresponds to an enzyme of the
XX CC invention.
XX SQ Sequence 421 AA;
Query Match 20.6%; Score 1248.5; DB 22; Length 421;
Best Local Similarity 65.2%; Pred.No. 6.9e-87;
Matches 251; Conservative 58; Mismatches 63; Indels 13; Gaps 3;
QY 749 VQYGGSSVADARRIRRAEIRIVATKKQGNVWVVSAMGDTTDDLDDLAQQVCPADPPR 808
Db 5 VQYGGSSLESARRIRRAEIRIVATKAGNDVVVCSAMGDTTDELDLAAAVNVPVPPAR 64
QY 809 ELDMLITAGERISNALVMAIESLGAHARSFTGSAQVITGTGHNAKIIDVTFGRLOTA 868
Db 65 EMDMLITAGERISNALVMAIESLGAESAQFTGSAQVLTTRHGNARIVDVTFGRVREA 124
QY 869 LEGRVVLVYAGFCGVGSODTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 928
Db 125 LDSEKICIVAGFCGVNKETDVTTLGRGSGDTTAVALLAALGADVCEIYSDVGVYVADP 184
QY 929 RIVRNARKLDTVTFEEMLEMAACAQVLMRLCRVEYARRHNI PVHVRSSYSYDRPGTVVVG 988
Db 185 RIVPNAQKLERLSFEEMLEAAVGSKILVLSVEYARAFVPMVRVRSYSNDPQTLLIAGS 244
QY 989 IKDVPMDPILTVGAHDRSAKTVIVGLPDI PGVAAKVFAVARRR-----RQHRIGAA 1042
Db 245 MEDIPMEEAVLTGVATDKSEAKVTVLGIPDKPGAAXVFRALADAEINIDMLVQNVSSE 304
QY 1043 ERLQGRGRQDRHLLHLLPQTSQGPVPPKKNWTRSETRSASTOLLYDDHIGKVSILIGAGMRSH 1102
Db 305 D-----GTTD--ITFTCPRSDGPRAMELLKKMQQGGDTWVLVDQGVKVSILVAGMKSH 357
QY 1103 PGVTATFCEALAAVGVNIELISTSE 1127
Db 358 PGVTAFMEALRDVNVNVELISTSE 382
RESULT 12
AAR63568
ID AAR63568 standard; Protein; 842 AA.
XX
XX AAR63568;
XX DT
XX DT
XX DT
XX DT
XX Feedback inhibition release aspartokinase 279Thr, 301Ser, 308Thr.
XX Feedback inhibition release aspartokinase; Brevibacterium flavum;
KW Corynebacterium; L-lysine production.

```

[illegible]

QY 803 PAPPRELDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNKIIDVTP 862
DB 480 PVPAREMDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNKIIDVTP 539
QY 863 GRLOTALEGRVVLVAGFQVSDTDXDVTTLGRGSDTTAVAMAAALGADVCEIYTDVVG 922
DB 540 GRVREALDEGKICIVAGFQVGNKTRDVTTLGRGSDTTAVAMAAALGADVCEIYTDVVG 599
QY 923 IFSADPRIVRNARKLDTVTPEEMLEMAACGAKVLMURCVYARHNIPIVHVRSSYSRDPG 982
DB 600 VYTADPRIVPNAQKLEKLSFEEMLEMAACGAKVLMURCVYARHNIPIVHVRSSYSRDPG 659
QY 983 TVVVGSIKDVPMEDPILTVGAHRSSEAKVTIVGLPDIPGYAAKVRAVA-----R 1032
DB 660 TLIAGSMEDIPVEEAVLTGVATDKSEAKVTIVGLSDXPGTAKVFRALADAEINIDMWLQ 719
QY 1033 RRQHRHGAERLQGRQD-RHHLHLLP--QTSGPPPMKWNTRSETRSASTOLLYDDHI 1089
DB 720 NVSSVEDGTTDITFTCPRSDDRAMEILKKLVQV-----NW-----TNVLYDDQV 765
QY 1090 GKUSLIGAGMRSHPGVTATFCEALAAVGNIELISTSE 1127
DB 766 GKUSLVGAGMKSHPGVTAEFMEALRDVNNVNIELISTSE 803
RESULT 14
AAR63569
ID AAR63569 standard; Protein; 842 AA.
AC AAR63569;
XX
XX 29-JUN-1995 (first entry)
XX Feedback inhibition release aspartokinase 279Val, 301Ser, 308Thr.
DE
XX Feedback inhibition release aspartokinase; Brevibacterium flavum;
KW Corynebacterium; L-lysine production.
XX Brevibacterium flavum.
XX JP06261766-A.
XX
XX 20-SEP-1994.
XX
XX 16-MAR-1993; 93JP-0055451.
XX
XX 16-MAR-1993; 93JP-0055451.
XX (MTP) MITSUBISHI PETROCHEMICAL CO LTD.
XX WPI; 1994-337417/42.
DR N-PSDB; AAQ72695.
XX
XX DNA encoding feedback inhibition release asparto-kinase - used
PT for the production of L-lysine.
XX
XX Claim 4; p19; 28pp; Japanese.
XX
XX AAQ72695 encodes AAR63569 Brevibacterium flavum feedback inhibition
CC release aspartokinase, it was used in the construction of a
CC plasmid. The plasmid was used to transform Corynebacterium cells,
CC where the expression of AAQ72695 resulted in increased L-lysine
CC production.
XX
XX Sequence 842 AA;
XX
XX Query Match 20.6%; Score 1248; DB 15; Length 842;
XX Best Local Similarity 36.8%; Pred. No. 2.1e-86;
XX Matches 345; Conservative 129; Mismatches 241; Indels 224; Gaps 31;
XX
XX 255 QVCTTRBAGYKDVILBSAITTIWDFSDVAADKVLGYRNWGLNKGDIAAAADKDG 314
DB 25 RIVATKAGNNVVVCSANG-----DITDELL-----ELAAAVNP-- 59

QY 315 TAFILRVLRDRNYTAPCGGQFTLPGBSLMEFVRNVGHLMTNDAIVDTDGGSEVPEGIMDALF 374
DB 60 -----VPPAREMDMLLTAG-----ERISNALV 81
QY 375 TGLIAIHGLKASDVNGPLINSRGTSGIYIVKPKMHGPAEVAFTCELSFVEDVIL----- 427
DB 82 AMATESLGAERASFTG-----SQAG-----VLTTERHGNARIVDVTTP--GRVREALDEGKICI 132
QY 428 -----GLPQNTMKIGIMDE--BRRTTVNLKACIKAK-----AADRUVVFINTGF 457
DB 133 VAGFQVGNKTRDVTTLGRGSDTTAVAMAAALNADVCEIYSDVGVYTAADPRIVPNAQK 192
QY 468 LDRTGDSIIHTSMEA-GPMVRKGTMKSQPMTLAY-----EDHNVDAG--LAAGFSGRAQV 518
DB 193 LEKLSFEEMLELAAGVSKIL--VIRSVYARAFNVPLRVRSYSNPDPTLIAGSMEDIPV 250
QY 519 GKGMWMTMELMADVETKIAQPRAGASTAVPSPPTAATLHALHYQVDVAAVQGLAGKR 578
DB 251 EEA--LTGVATDKSEAKVTV--LGISDK--PGEVAKVFRALADAEINIDMWLQNVSSVE 304
QY 579 RATIEQLLTIP-----LAKELAWAPDEIRREVDNQCOSILGYVVRWVDO----- 622
DB 305 DGTITDITFTCPRSDDRAMEILKKL-----QVGNWNTNL-----YDDQVGVKSLV 350
QY 623 GVCGSKVPDIHDVALMEDRATLRISQLLANLRLHGVTISADVRASLERMAPLVDQNAG 682
DB 351 GAGVKSHPGV-TAEFMEALRDVNNVIEL-----ISTSEIRIS-----VLIREDDL 394
QY 683 DVAYRMAPNPFDDSIAPLAAQELILSCAQPNGYTEPILHRRRREPKARAEAPSDRA 742
DB 395 DAAARALHEQF-----QLGGEDEAVVYA-----GT 419
QY 743 GDDAARYQKYGGSSVADAERIRRVAAERIVATKQGDVVDVVVVSAMGDTTDLDDLQAVC 802
DB 420 GRVALVTQKYGGSSLESAERIRNVAAERIVATKAGNNVVVCSAMGDTTDELLELAAYN 479
QY 803 PAPPRELDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNKIIDVTP 862
DB 480 PVPAREMDMLLTAGERISNALVAMAEISLGAHARSFTGSOAGVITTTGTHGNKIIDVTP 539
QY 863 GRLOTALEGRVVLVAGFQVSDTDXDVTTLGRGSDTTAVAMAAALGADVCEIYTDVVG 922
DB 540 GRVREALDEGKICIVAGFQVGNKTRDVTTLGRGSDTTAVAMAAALNADVCEIYSDVVG 599
QY 923 IFSADPRIVRNARKLDTVTPEEMLEMAACGAKVLMURCVYARHNIPIVHVRSSYSRDPG 982
DB 600 VYTADPRIVPNAQKLEKLSFEEMLEMAACGAKVLMURCVYARHNIPIVHVRSSYSRDPG 659
QY 983 TVVVGSIKDVPMEDPILTVGAHRSSEAKVTIVGLPDIPGYAAKVRAVA-----R 1032
DB 660 TLIAGSMEDIPVEEAVLTGVATDKSEAKVTIVGLSDXPGTAKVFRALADAEINIDMWLQ 719
QY 1033 RRQHRHGAERLQGRQD-RHHLHLLP--QTSGPPPMKWNTRSETRSASTOLLYDDHI 1089
DB 720 NVSSVEDGTTDITFTCPRSDDRAMEILKKLVQV-----NW-----TNVLYDDQV 765
QY 1090 GKUSLIGAGMRSHPGVTATFCEALAAVGNIELISTSE 1127
DB 766 GKUSLVGAGMKSHPGVTAEFMEALRDVNNVNIELISTSE 803
RESULT 14
AAR63569
ID AAR63569 standard; Protein; 421 AA.
XX
XX AAR63569;
XX
XX 19-FEB-2001 (first entry)
XX
XX Modified Corynebacterium glutamicum aspartokinase, T3111.
XX Aspartokinase; AK; lysine synthesis; feedback inhibition;
KW

[illegible]

Db	305	DGTTDIIIFTCPRSDGRRRAWEIKKLQVQG-----NW-----TNNVLYDDQGVKSLV	350
Qy	1096	GAGMRSHPGVTAIFCEALAAAGVNNIELISTSE	1127
Db	351	GAGMKSHPGVTAIEFWEALRDVNNVIELISTSE	382
RESULT 15			
ID	ABP97764		
AC	ID ABP97764 standard; Protein; 421 AA.		
XX	ABP97764;		
DT	28-MAY-2003 (first entry)		
XX	Amino acid sequence of an aspartate kinase carrying T311I mutation.		
DE	Corynebacterium bacteria; chemical compound; L-amino acid; vitamin; lysC;		
KW	aspartate kinase.		
XX	Synthetic.		
OS	Corynebacterium glutamicum.		
XX	OS		
FH	Key Location/Qualifiers		
FT	Misc-difference 1		
FT	/note= "Met encoded by GTG"		
FT	Misc-difference 311		
FT	/note= "wild type Thr changed to Ile"		
XX	WO2003014330-A2.		
XX	20-FEB-2003.		
XX	30-JUL-2002; 2002WO-EP08465.		
PF	06-AUG-2001; 2001US-309877P.		
XX	(DEGS) DEGUSSA AG.		
XX	Bathe B, Kreutzer C, Moeckel B, Thierbach G;		
XX	WPI; 2003-278476/27.		
DR	N-PSDB; ABZ77392.		
XX	Coryneform bacteria to produce chemical compounds e.g. L-lysine, has at		
PT	least two copies of open reading frame, gene or allele in question at a		
PT	particular desired site		
XX	Claim 16; Page 105-106; 109pp; English.		
XX	The specification describes Coryneform bacteria which produce chemical		
CC	compounds, which instead of a single copy of an open reading frame (ORF),		
CC	gene or allele naturally present at a desired locus have at least two		
CC	copies of the ORF, gene or allele, preferably in tandem arrangement.		
CC	The modified bacteria are useful for producing chemical compounds		
CC	including L-amino acids, vitamins, nucleosides and nucleotides. The		
CC	present sequence is encoded by a lysC gene, and represents an aspartate		
CC	kinase carrying the T311I mutation. Tandem duplications of the lysC		
CC	gene are used to construct the bacteria of the invention.		

XX	SQ	Sequence	421 AA;
		Query Match	20.3%; Score 1227.5; DB 24; Length 421;
		Best Local Similarity	64.08; Pred No. 2.8e-85;
		Matches 251; Conservative	56; Mismatches 58; Indels 27; Gaps 5;
QY		749 VQKYGGSSVADAERIRRVVAERI	VATKFKGNDVVVVVSAMGDTTDDLLDLAQOVCPPPPR 808
		:	:
		:	:
		:	:
DB		5 VQKYGGSSLSAEIRNVAERI	VATKAGNDVVVVCSAMGDTTDELLEALAAVNVPVPPAR 64
QY		809 ELDMLLTAGERISNALVMAAIESLG	AHARSFTGSQAQVITTTGTGHNAKIIDVTTPGRLQTA 868
		:	:
		:	:
DB		65 EMDMLLTAGERISNALVMAAIESLG	AQAQSFSTGSAQLTTERIGNARIVDVTPGVREA 124
		:	:
		:	:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 24.7156 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-52
Perfect score: 6061
Sequence: 1 MQHHHHHTDRVSVGNLRIA.....SAATRRPCTGRGRWACQ 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	62.9	741	3	US-09-001-984C-106
2	3810	62.9	741	4	US-09-396-347P-106
3	2570	42.4	731	4	US-09-252-991A-22524
4	2389	37.8	739	4	US-08-836-943-2
5	2225	36.7	729	4	US-09-328-352-7603
6	1545.5	25.5	421	4	US-08-311-731A-30
7	1223.5	20.2	421	1	US-08-532-828B-3
8	1223.5	20.2	421	1	US-08-700-359-9
9	1223.5	20.2	421	2	US-08-596-366-6
10	1223.5	20.2	421	2	US-08-967-104-6
11	1223.5	20.2	421	3	US-08-985-908-5
12	1223.5	20.2	421	3	US-08-852-730-14
13	1223.5	20.2	421	3	US-08-985-916-5
14	1219.5	20.1	421	1	US-08-532-828B-4
15	827	13.6	461	4	US-07-254-991A-23720
16	785	13.0	411	1	US-07-684-135A-2
17	770	12.7	156	4	US-09-056-556-227
18	770	12.7	156	4	US-09-072-596-222
19	746.5	12.3	438	4	US-09-328-352-6511
20	563	9.3	424	4	US-09-134-001C-5225
21	364.5	6.0	820	2	US-08-380-182-23
22	314.5	5.2	262	4	US-08-887-534A-23
23	314.5	5.2	262	4	US-09-527-431-23
24	313.5	5.2	863	2	US-08-380-182-19
25	310.5	5.1	449	1	US-08-256-136-2
26	310.5	5.1	449	1	US-08-950-737-2
27	310.5	5.1	449	2	US-08-973-461A-8

28	310.5	5.1	449	3	US-08-648-010-8	Sequence 8, Appli
29	310.5	5.1	863	2	US-08-380-182-20	Sequence 20, Appl
30	294.5	4.9	463	4	US-09-134-001C-4160	Sequence 4160, Ap
31	285.5	4.7	172	1	US-08-532-828B-5	Sequence 5, Appli
32	285.5	4.7	172	1	US-08-700-359-11	Sequence 11, Appl
33	285.5	4.7	172	3	US-08-985-908-7	Sequence 7, Appli
34	285.5	4.7	172	3	US-08-852-730-16	Sequence 16, Appl
35	285.5	4.7	172	3	US-08-985-916-7	Sequence 7, Appli
36	281.5	4.6	172	1	US-08-532-828B-6	Sequence 6, Appli
37	281.5	4.6	172	2	US-08-596-366-8	Sequence 8, Appli
38	281.5	4.6	172	2	US-08-967-104-8	Sequence 8, Appli
39	241	4.0	441	4	US-09-198-452A-1124	Sequence 1124, Ap
40	226	3.7	478	4	US-09-107-532A-6090	Sequence 6090, Ap
41	181.5	3.0	3892	4	US-09-328-352-5503	Sequence 5503, Ap
42	172.5	2.8	4545	2	US-08-804-227C-14	Sequence 14, Appl
43	172.5	2.8	4550	2	US-08-804-227C-8	Sequence 8, Appli
44	172.5	2.8	4550	2	US-08-804-198-2	Sequence 2, Appli
45	171	2.8	3798	3	US-09-335-409-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-001-984C-106
; Sequence 106, Application US/09001984C
; Patent No. 6245331
; GENERAL INFORMATION:
; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Belisle, John T
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: NYU-011
; CURRENT APPLICATION NUMBER: US/09/001,984C
; PRIOR FILING DATE: 1997-12-31
; PRIOR APPLICATION NUMBER: 60/034,003
; PRIOR FILING DATE: 1996-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-001-984C-106

Query Match	62.9%	Score	3810	DB	3	Length	741
Best Local Similarity	100.0%	Pred. No.	0				
Matches	740	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	9	TDVSVGNLRIARVLVDFVNNALPGTDIDPDSFWAGVDKVVADLTPQNALNARDELQ	68				
Db	2	TDVSVGNLRIARVLVDFVNNALPGTDIDPDSFWAGVDKVVADLTPQNALNARDELQ	61				
Qy	69	AQIDKWHRRRVIPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDABITTTAGPQLVVPV	128				
Db	62	AQIDKWHRRRVIPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDABITTTAGPQLVVPV	121				
Qy	129	LNARFALNANARWGLYDALYGTDPVETDGAKEGPTYNKVRGDKVIARFKFLDSDVP	188				
Db	122	LNARFALNANARWGLYDALYGTDPVETDGAKEGPTYNKVRGDKVIARFKFLDSDVP	181				
Qy	189	LSSGSFGDAGTFTVDGQLVVALPKSTGLANPQFAGYTGAAESPTSVLLINHLHIEI	248				
Db	182	LSSGSFGDAGTFTVDGQLVVALPKSTGLANPQFAGYTGAAESPTSVLLINHLHIEI	241				
Qy	249	LIDPESQVGTDRAGVKDVIRESAITTIMDFESVAADKVLGYRNWLGKNGDLAA	308				
Db	242	LIDPESQVGTDRAGVKDVIRESAITTIMDFESVAADKVLGYRNWLGKNGDLAA	301				
Qy	309	AVDCKGTAFLRVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHMTNDALVDTGSEVFEG	368				
Db	302	AVDCKGTAFLRVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHMTNDALVDTGSEVFEG	361				

369 QY IMDALFTGLIAIRGLKASDVNGPLINSRTGSIYIVKPKHGPAAEVAFTCELSRVEDVLG 428
362 DB IMDALFTGLIAIRGLKASDVNGPLINSRTGSIYIVKPKHGPAAEVAFTCELSRVEDVLG 421
429 QY LPQNTWKIGIMDBERTTNLAKACIKAAADRVVFIINTGFLDRTGDEIHTSMAGPMVRKG 488
422 DB LPQNTWKIGIMDBERTTNLAKACIKAAADRVVFIINTGFLDRTGDEIHTSMAGPMVRKG 481
489 QY TMSQPMILAYEDHNVNDAAGSFGSAQVKGKGMWMTLMADMVETKIAQPRAGASTAW 548
482 DB TMSQPMILAYEDHNVNDAAGSFGSAQVKGKGMWMTLMADMVETKIAQPRAGASTAW 541
549 QY VPSPTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRREVDNN 608
542 DB VPSPTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRREVDNN 601
609 QY CQSILGYVVRWVDOGVGCKSKVPDIHVALMEDRATLRISSOLLANWLHGVITTSADVRAS 668
602 DB CQSILGYVVRWVDOGVGCKSKVPDIHVALMEDRATLRISSOLLANWLHGVITTSADVRAS 661
669 QY LERMAPLVDRONAGDVAYRPMAPNFDDSIATFLAAQELILSGAQPNNGYTEPILHRRREF 728
662 DB LERMAPLVDRONAGDVAYRPMAPNFDDSIATFLAAQELILSGAQPNNGYTEPILHRRREF 721
729 QY KARAEEKPAPSDRAGDDAAR 748
722 DB KARAEEKPAPSDRAGDDAAR 741

RESULT 2
US-09-396-347F-106
; Sequence 106, Application US/09396347F
; Patent No. 6506384
; GENERAL INFORMATION:
; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Belisle, John T
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: 32004-169276
; CURRENT APPLICATION NUMBER: US/09/396,347F
; PRIOR FILING DATE: 1999-09-14
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-396-347F-106

Query Match 62.9%; Score 3810; DB 4; Length 741;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

9 TDRVSVGNLRIARVLYDFVNNALPGTIDIDPSFWAGVGVKVVADLTPOQALLNARDELQ 68
2 TDRVSVGNLRIARVLYDFVNNALPGTIDIDPSFWAGVGVKVVADLTPOQALLNARDELQ 61
69 AQIDKXWRRVIEPIQMDAYRQPLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVWPV 128
62 AQIDKXWRRVIEPIQMDAYRQPLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVWPV 121
129 LNARFALNANARWGSILYDALYGTVDVIPETDGAEGKPTYNKVRGDKVIAYAKFLDDSDVP 188
122 LNARFALNANARWGSILYDALYGTVDVIPETDGAEGKPTYNKVRGDKVIAYAKFLDDSDVP 181
189 LSSGSGDATGFTVQDQQLVWALPKSTGLANPGQFAGYTGAAESPTSVLLINHLGHEI 248
182 LSSGSGDATGFTVQDQQLVWALPKSTGLANPGQFAGYTGAAESPTSVLLINHLGHEI 241
249 LIDPESQVGTTRAGVKDVLIESAITTTIMDFEDSVAAVDAADKVLGYRNWGLNKGDLAA 308

242 DB LIDPESQVGTTRAGVKDVLIESAITTTIMDFEDSVAAVDAADKVLGYRNWGLNKGDLAA 301
309 QY AVDXDGTAFLEVLNRDRNYTAPGGQFTLFGSRSLMFVRNVGHLMNTNDIAIVDTGSEVFEG 368
302 DB AVDXDGTAFLEVLNRDRNYTAPGGQFTLFGSRSLMFVRNVGHLMNTNDIAIVDTGSEVFEG 361
369 QY IMDALFTGLIAIRGLKASDVNGPLINSRTGSIYIVKPKHGPAAEVAFTCELSRVEDVLG 428
362 DB IMDALFTGLIAIRGLKASDVNGPLINSRTGSIYIVKPKHGPAAEVAFTCELSRVEDVLG 421
429 QY LPQNTWKIGIMDBERTTNLAKACIKAAADRVVFIINTGFLDRTGDEIHTSMAGPMVRKG 488
422 DB LPQNTWKIGIMDBERTTNLAKACIKAAADRVVFIINTGFLDRTGDEIHTSMAGPMVRKG 481
489 QY TMSQPMILAYEDHNVNDAAGSFGSAQVKGKGMWMTLMADMVETKIAQPRAGASTAW 548
482 DB TMSQPMILAYEDHNVNDAAGSFGSAQVKGKGMWMTLMADMVETKIAQPRAGASTAW 541
549 QY VPSPTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRREVDNN 608
542 DB VPSPTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRREVDNN 601
609 QY CQSILGYVVRWVDOGVGCKSKVPDIHVALMEDRATLRISSOLLANWLHGVITTSADVRAS 668
602 DB CQSILGYVVRWVDOGVGCKSKVPDIHVALMEDRATLRISSOLLANWLHGVITTSADVRAS 661
669 QY LERMAPLVDRONAGDVAYRPMAPNFDDSIATFLAAQELILSGAQPNNGYTEPILHRRREF 728
662 DB LERMAPLVDRONAGDVAYRPMAPNFDDSIATFLAAQELILSGAQPNNGYTEPILHRRREF 721
729 QY KARAEEKPAPSDRAGDDAAR 748
722 DB KARAEEKPAPSDRAGDDAAR 741

RESULT 3
US-09-252-991A-22524
; Sequence 22524, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22524
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22524

Query Match 42.4%; Score 2570; DB 4; Length 731;
Best Local Similarity 69.1%; Pred. No. 9.6e-222;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;

9 TDRVSVGNLRIARVLYDFVNNALPGTIDIDPSFWAGVGVKVVADLTPOQALLNARDELQ 68
8 TERVQVGLQVAKVDFVNNALPGTGVSAIDFTWGAEAVINDLAPKNKALLAKXDELQ 67
69 AQIDKXWRRVIEPIQMDAYRQPLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVWPV 128
68 AKIDGWHQARAGQAHAAYAKFLEEIGYLLPEADFOAGTQNVDDDEIARMAGPQLVWPV 127
129 LNARFALNANARWGSILYDALYGTVDVIPETDGAEGKPTYNKVRGDKVIAYAKFLDDSDVP 188
128 MNARFALNANARWGSILYDALYGTVDVISEGGKKGKYNKVRGDKVIAYAKFLDDSDVP 187

189 LSSGSGDATGTTVODGQVVALPKS--TGLANPCQFAGYTCAAESPTSVLLINHGHLIE 247
189 LESGSHVATSVSVKNGALVWALKNGSETGLKNAQGFQDAKFAQVLLKINGLHFE 247
248 ILIDPESQVGTDRAGVQVIVLESNITIMPFESVAAVDAADKVLGYRNWLGKNGDLA 307
248 IQIDSSSVGQDAAAGVNDVLEALITIMPCDSVAAVDAADKVLGYRNWLGKNGDLA 307
308 AAVDGDGTAFRLVRNDRNYTAPGGQFTLPGRSLMFVRNVLHMTNDIAVDTDGSEVFE 367
308 BEVSKGGSTFTTMMPDVYTRADGSELTGHSRLLFVRNVGHMTNDAILDKGNEVPE 367
368 GIMDALFTGLIAIHGLKASDVNGPLI--NSRTGSTYIVYKPMHGPFAEVAFTCELFERSVED 425
368 GIDGLFTSLAIH-----DLNGNTRSRKNSRTGSYIVYKPMHGPFAEVAFTCELFERSVED 422
426 VLGLPONTMKGIMDEBERTVNLKACIKAAADRVVFNITGFLDRTGDEIHTSWEAGPMV 485
423 VLGLPNTLVKGMDEBERTVNLKACIKAAADRVVFNITGFLDRTGDEIHTSWEAGAVV 482
486 RKGTMKSQFWILAYEDHNVDAGLAAGFSGRAQVKGMMTMTLMADVMVETXIAQPRAGAS 545
483 RKGAMKSEKIGAVENNVVGLATGLQKQAIQKGMWAMPDLMAAMLEQKIGHPLAGAN 542
546 TAWVPSPTAATHALHYHOVDVAAVQOGLAGKRRATIEQLITIPLAKELAWAPDEIRREV 605
543 TAWVPSPTAATHALHYHKVDVFARQAEIAKETPASVDDIITIPLAPNTNWTABEIKNEV 602
606 DNNCOSILGYVVRVWDQVGGSKVPDIHDVALMEDRATLRSSOLLANLWLRHGHVITGADV 665
603 DNNAGILGYVVRVWDQVGGSKVPDIHDVALMEDRATLRSSOLLANLWLRHGHVITGADV 662
666 RASLERMAPLVDQRNAGVAVRPMAPNPDSDIAFLAQELILSGAQOQNGYTEPILHRRR 725
663 VESLKRMAVVVDQRNASDPSYRPMAPNPDSDIAFLAQELILSGAQOQNGYTEPILHRRR 722
726 REFRAK 731
723 REFRAK 728

RESULT 4

US-08-836-943-2
; Sequence 2, Application US/08836943
; Patent No. 5965391
; GENERAL INFORMATION:
; APPLICANT: Reinscheid, Dieter
; APPLICANT: Elkmanns, Bernhard
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: DNA WHICH REGULATES GENE EXPRESSION IN
; TITLE OF INVENTION: CORYNEFORM BACTERIA
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Firm of Karl F. Ross, PC
; STREET: 5676 Riverdale Ave.
; CITY: Bronx
; STATE: New York
; COUNTRY: USA
; ZIP: 10471
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,943
; FILING DATE: 08-MAY-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Jonathan
; REGISTRATION NUMBER: 26,963
; REFERENCE/DOCKET NUMBER: 20357
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (718) 884-6600
; TELEFAX: 718/601-1099
; TELEX: 620428
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 739 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-836-943-2

Query Match 37.8%; Score 2289; DB 2; Length 739;
Best Local Similarity 61.3%; Pred. No. 1.7e-196;
Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;
QY 9 TDRSVGNLRIRLVYDFVNNALPCTDIDPDSFWAGVDKVVADLTPOQALINAEDELO 68
DB 20 TEVDAGGQVAKVLDYFTEAVLPVGVDAKFGSFAIARDLTPRRELLARDELO 79
QY 69 AQIDKWHRRVIEPIDMDAYRQFLEIGYLLPEPDDFTITSGVDAEITTTAGPQLVVPV 128
DB 80 MLIDDYHRNN-SCITQEAIVEDFLKEIGYLVVEEPEAAEIRTQNVDTESISSTAGPQLVVP 138
QY 129 LNARFALNANARWGLSYDALYGTVDVLPETDGAERKPTNKVRGDKVIAYAKFLDSDVP 188
DB 139 LNARFALNANARWGLSYDALYGTNAIPETDGAERKPTNKVRGDKVIAYAKFLDSDVP 198
QY 189 LSSGSGDATGFTVODGQVVALPKDSTGLANPCQFAGYTCAAESPTSVLLINHGHLIE 248
DB 199 LDGASHADYVEKYNITDGLAAHIGDSVYRLKNRESYRGFTGNFLDPEALLLETINGHLIEL 258
QY 249 LIDPESQVGTDRAGVQVIVLESNITIMPFESVAAVDAADKVLGYRNWLGKNGDLA 308
DB 259 QIDPVHPIGKADTGLKDVLESNITIMPFESVAAVDAADKVLGYRNWLGKNGDLA 318
QY 309 AVDKDGTAFRLVRNDRNYTAPGGQFTLPGRSLMFVRNVLHMTNDIAVDTDGSEVFE 368
DB 319 EMSKNGRIFTRELKNDRVYIGRNGTELVHGRSLLFVRNVGHLMQNPDL-IDGEEIFEG 377
QY 369 IMDALFTGLIAIHGLKASDVNGPLINSRTGSYIVYKPMHGPFAEVAFTCELFERSVEDVLG 428
DB 378 IMDALFTVCAIPGIAPQN---KMRNSRKGSYIVYKPMHGPFAEVAFTCELFERSVEDL 434
QY 429 LPONTMKGIMDEBERTVNLKACIKAAADRVVFNITGFLDRTGDEIHTSWEAGPMV 488
DB 435 LPRHTLVGVNDEBERTSVNLDASINMEVADRLAFINTGFLDRTGDEIHTSWEAGPMV 494
QY 489 TMSQFWILAYEDHNVDAGLAAGFSGRAQVKGMMTMTLMADVMVETXIAQPRAGASTAW 548
DB 495 DMOTAPKQAYENNVNVDAGIQRGLPKCAQIGKGMWAMPDLMAAMLEKIKICOPREGANTAW 554
QY 549 VPSPTAATHALHYHOVDVAAVQOGL-AGKRRATIEQLITIPLAKELAWAPDEIRREV 607
DB 555 VPSPTGATLHATHYHLVDYFKVQDELRAAGRRDLENILTIPTAPNTNWESEKKEMDN 614
QY 608 NCOSILGYVVRVWDQVGGSKVPDIHDVALMEDRATLRSSOLLANLWLRHGHVITSDVRA 667
DB 615 NCOSILGYVVRVWEHGVGGSKVPDIHDVALMEDRATLRSSOLLANLWLRHGHVITSDVRA 674
QY 668 SLERMAPLVDQRNAGVAVRPMAPNPDSDIAFLAQELILSGAQOQNGYTEPILHRRR 727
DB 675 SLERMAVVVDQRNASDPSYRPMAPNPDSDIAFLAQELILSGAQOQNGYTEPILHRRR 734
QY 728 FKAR 731
DB 735 FKAR 738

RESULT 5

US-09-328-352-7603
; Sequence 7603, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7603
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7603

Query Match 36.7%; Score 2225; DB 4; Length 729;
Best Local Similarity 59.8%; Pred. No. 9.5e-191;
Matches 432; Conservative 97; Mismatches 185; Indels 8; Gaps 4;

QY 9 TDRVSGNRIARVLYDFVNEALPGTDIDPDSFWAGYDKVADVADLPQKALLNARDELQ 68
DB 10 TARIQKLAIAKELYDFIENEALPGSLDSEYKQVQVVDLSPKNAKLLAKRDELQ 69

QY 69 AQIDKWHRRVIRPIDWDAYKQFLTEIGVLLPEPDDFTTTSQVDAEITTTAGPOLVVPV 128
DB 70 AKIDEWHRRNKF--LGAYKAFLEIGVLLPEVEDFQITENVDEEIALLAGPOLVVPV 126

QY 129 LNARFALNANARWGLYDALYGTDPETDGAEGKPTYNKVRGDKVIAIYAKFLDDSV 188
DB 127 RNARYCLNANARWGLYDALYGFVISEEGGAEGKGYNPVRGAKEVFAKNFLNEIFP 186

QY 189 LSSGSGDATFTVQDQGVVLPALPD-KSTGLANPGQFAGYTGAAESPTSVLLINGLHIE 247
DB 187 LAQGSADATKYAIEQKLVLT-KDGTGTLAHEAQFVGFNGEENPFSEVLLSGLHVI 246

QY 248 ILIDPESQVTTDRAGVKDVLIESALTITMDPESVAAVDAADKVLGYRNWLGKNGDIA 307
DB 247 IEIDANSPIQTDLAGVKOLTLEAAVTTIQDLEDSVAAVDAEKEVGYENWLGKNGTLQ 306

QY 308 AAVDKGTAPLRVLRNDRNTPAGGQFTLPGRSLMFVRNVGHMTNDAIVTDDGSEVPE 367
DB 307 ESIEKNGKTIIVRALNKDREIKN-LIGGTTKLHGRSLMLLRNVGHMTNPAIL-VDGEEIPE 365

QY 368 GIMDALFTGLIATHGLKASDVNGPLNSRTGSIYIVKPKVHGPAEVAFTCELFPSVEDVL 427
DB 366 GIMDALVTPLLSLADIRSENEK--NSRKGSMYIVKPKVHGPEEVAFAVELFERAEQAL 422

QY 428 GLPONTWKIGIMDEERTTNLAKACIAAADRVVINTGFLDRTGDEIHTSMEAGPMVRK 487
DB 423 GLPAKSLKIGIMDEERTSNLKNCAIAAAKDRITFINTGFMDBRTGDEIHTSMEAPVVRK 482

QY 488 GTWKSQPWILAYEDHNVDAAGAFSGRAQVKGKMTMTLMADMVETKIAOPRAGASTA 547
DB 483 EAVKTOKWIAAYENRNVAIGKCGLOGKQAGIKGKMPKPSMKMDLMTAAAPNAGASCA 542

QY 548 WVPSPATATLHALHYHQVDVAAVQOGLAGKRRTATIEQLLTIPLAKELAWAPDEIEEVDN 607
DB 543 WVPSPGAVLHAMHYHQVNVKAAQDQKAEEMLSLDLLTTPFATDTNWSAEIENNELN 602

QY 608 NCQSILGVYRWVDQGVGSKVPDIDHDVALMEDRATIRISSOLLANLWLRHGVITADVPA 667
DB 603 NCQGIILGVYRWVDLGVGSKVPDINNVMGDMEDRATIRISSQHVANLWLRHGVITREQVE 662

QY 668 SLERMAFLVRQAGVAVRPMAPNPFDSIAFLAAQELISGAQOQNGVTEPILHRRRRE 727
DB 663 VLKRMKIVDEQANNDPLYKPMANFETNIAFOAASDLIFKCEQPSGYTEPLLHAARLK 722

QY 728 FX 729
DB 723 LX 724

RESULT 6
US-08-311-731A-30
; Sequence 30, Application US/08311731A

; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM LEPRAE
US-08-311-731A-30

Query Match 25.5%; Score 1545.5; DB 4; Length 421;
Best Local Similarity 80.4%; Pred. No. 4.2e-130;
Matches 319; Conservative 18; Mismatches 43; Indels 17; Gaps 3;

QY 749 VQYGGSSVADAERIRRAERIVATKQGNVWVWSAMGDTTDDLDDLAQQVCPPPPR 808
DB 5 VQYGGSSVADADRIRRAERIVQTKQGNDIVVWSAMGDTTDDLDDLAQQVCPPPPAR 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTQSGAVITTTGTHGNAKIIDVTPLQTA 868
DB 65 ELDMLLTAGERISNALVAMAIESFGAARSFTQSGAVITTTGTHGNAKIIDVTPLQTA 124

QY 869 LEEGRVVLVAGFQGVSDTQKVTTLGRGGSDTTAVAVAAALGADVCEIYTDVDFISADP 928
DB 125 LDEGRVVLVAGFQGVSDTQKVTTLGRGGSDTTAVAVAAALGADVCEIYTDVDFISADP 184

QY 929 RIVNARKLDTVFEEMLEMAACGAKVLMRCVYARRHNIPIVHVRSSISDRGTGVVGS 988
DB 185 RVFNARKLDTVFEEMLEMAACGAKVLMRCVYARRHNIPIVHVRSSISDRGTGVVGS 244

QY 989 IKDVPMDPILTVGAHDSERAKVTIVGLPDI PGYAAKVFAVARRRRQRHGAEE----- 1043
DB 245 IKDVPMDPILTVGAHDSERAKVTIVGLPDI PGYAAKVFAVARRRRQRHGAEE-----DADVNIDM 295

QY 1044 RLQGRGQDRHHGLH-----LPQSGPPPKWNTSETRSATQLLYDDHICKVSLIGAGMR 1100
DB 296 VLQNVSKVEDGKTDTITTCSDSGPIAVAKGLSLRDEIGFTQLLYDDHICKVSLIGAGMR 355

QY 1101 SHPGVTATFCEALAAVGVNIELISTSEDSRSCCAAT 1137
DB 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDT 392


```

RESULT 7
US-08-532-828B-3
; Sequence 3, Application US/08532828B
; Patent No. 5688671
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: OGAWA, Yuri
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: TANAKA, Akiko
; APPLICANT: MATSUI, Hiroshi
; TITLE OF INVENTION: MUTANT ASPARTOKINASE GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS TEXT EDITOR
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,828B
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-101450
; FILING DATE: 27-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-764-0 PCT
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC13869
; US-08-532-828B-3

Query Match 20.2%; Score 1223.5; DB 1; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

Qy 749 VQYGGSSVADAEIRIRVAERIVATKKGNDVVVVSAMGDTTDDLDAQQVCPAPPPR 808
Db 5 VQYGGSSLESAERIRNVAERIVATKAGNDVVVCSAMGDTTDELLEAAVNPVPPAR 64
Qy 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAQVITTTGTHGNAKIIDVTPGRLOTA 868
Db 65 EMDMLLTAGERISNALVAMAIESLGAESAQSFSGAQVLTTERHGNARIVDVTGVRVREA 124
Qy 869 LEEGRVVLVAGQGVSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 928
Db 125 LDESKI CIVAGFQGVNKETRVTTILGRGSDTTAVAAALNADVCEIYSDVDGVYADP 184
Qy 929 RIVNARKLDTVTTEMLEMAACGAKVLMLEFCVEYARRHNIPVHVRSSYSRDPGTVVVGS 988
Db 185 RIVENAQKLEKLSPEEMLELAAGSKILVLRSEYARAFNVFLRVRRSSYNDFGLIAGS 244

```

```

Qy 989 IKDVPMEDPILTCVAHDRSEAKVTIVGLPDIPGYAAKVERAVA-----RRRRQHR 1038
Db 245 MEDIPVEEAVLTGATDKSEAKVTIVGLSDKPGEAARKVFRALADAEINIDMWLQNVSSVE 304
Qy 1039 HGAERLQGRGROD-RHHLHLP--QTSPPPPKWNWTRSETRSASTQLLYDDHIGKVSLL 1095
Db 305 DGTDTITFTCTPRADGRAMEILKQLVQG-----TNVLYDDQGVKVSLLV 350
Qy 1096 GAGMRSHPGVGTATFCEALAAVGNIELISTSE 1127
Db 351 GAGMKSHPGVGTAEFMEALRDVNVNIELISTSE 382

RESULT 8
US-08-700-359-9
; Sequence 9, Application US/08700359
; Patent No. 5766925
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, MASAKAZU
; APPLICANT: USUDA, YOSHIHIRO
; APPLICANT: SUZUKI, TOMOKO
; APPLICANT: TANAKA, AKIKO
; APPLICANT: MATSUI, HIROSHI
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS TEXT EDITOR
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,359
; FILING DATE: 08-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-35019
; FILING DATE: 04-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-819-0 PCT
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-700-359-9

Query Match 20.2%; Score 1223.5; DB 1; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

Qy 749 VQYGGSSVADAEIRIRVAERIVATKKGNDVVVVSAMGDTTDDLDAQQVCPAPPPR 808
Db 5 VQYGGSSLESAERIRNVAERIVATKAGNDVVVCSAMGDTTDELLEAAVNPVPPAR 64
Qy 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAQVITTTGTHGNAKIIDVTPGRLOTA 868
Db 65 EMDMLLTAGERISNALVAMAIESLGAESAQSFSGAQVLTTERHGNARIVDVTGVRVREA 124
Qy 869 LEEGRVVLVAGQGVSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 928

```

Db 125 LDEGKICIVAGFGVGNKETRDVITLGRGSDTTAVALAAALNADVCEIYSDVDGVYTADP 184
QY 929 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVGVS 988
Db 185 RIVPNAOKLEKLSFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVGVS 244
QY 989 IKDVPMEDEPILGTVAHDSREAKVTIVGLPDIPGYAAKVFRAVA-----RRRRQHR 1038
Db 245 MEDIPVEEAVLTGVAIDKSEAKVTIVGLSDKPGCEAAKVFRALADAEINIDMWLQNVSSVE 304
QY 1039 HGAERLQGRQD-RHHLHLP--QTSPPPKWNTSRSETSASTOLLYDHIGKVSLLI 1095
Db 305 DGTDTITFTCPADGRAMEILKKLVQVQ-----TNVLYDDQGVKVSLLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGNVIELISTSE 1127
Db 351 GAGMKSHPGVTATFCEALAAVGNVIELISTSE 382

RESULT 9

US-08-596-366-6
; Sequence 6, Application US/08596366
; Patent No. 5876983
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHENOLPYRUVATE CARBOXYLASE,
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/596,366
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209775
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209776
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-153876
; FILING DATE: 05-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-784-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-596-366-6

Query Match 20.2%; Score 1223.5; DB 2; Length 421;
Best Local Similarity 64.0%; Pred. No. 3 6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 749 VQYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLLDLAQVCPAPPPR 808
Db 5 VQYGGSSLESASRIENVAERIVATKKAGNDVVVVCAMGDTTDELELAANVPVPPAR 64
QY 809 ELMWLTAGERISNALVAVAIESLGAHBSFTGSOAGVITTTGTHGNAKIIDVTGSLQTA 868
Db 65 EMDMLTAGERISNALVAVAIESLGAHBSFTGSOAGVITTTGTHGNARIVDVTGPEVRA 124
QY 869 LEEGRVVLVAGFGVGSQDTRDVTTLGRGSDTTTAVAMAALGADVCEIYTDVDFISADP 928
Db 125 LDEGKICIVAGFGVGNKETRDVITLGRGSDTTTAVALAALNADVCEIYSDVDGVYTADP 184
QY 929 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVGVS 988
Db 185 RIVPNAOKLEKLSFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVGVS 244
QY 989 IKDVPMEDEPILGTVAHDSREAKVTIVGLPDIPGYAAKVFRAVA-----RRRRQHR 1038
Db 245 MEDIPVEEAVLTGVAIDKSEAKVTIVGLSDKPGCEAAKVFRALADAEINIDMWLQNVSSVE 304
QY 1039 HGAERLQGRQD-RHHLHLP--QTSPPPKWNTSRSETSASTOLLYDHIGKVSLLI 1095
Db 305 DGTDTITFTCPADGRAMEILKKLVQVQ-----TNVLYDDQGVKVSLLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGNVIELISTSE 1127
Db 351 GAGMKSHPGVTATFCEALAAVGNVIELISTSE 382

RESULT 10
US-08-967-104-6
; Sequence 6, Application US/08967104
; Patent No. 5919694
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHENOLPYRUVATE CARBOXYLASE,
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,104
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/596,366
; FILING DATE: 29-APR-1996
; APPLICATION NUMBER: JP 5-209775
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209776
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-153876
; FILING DATE: 05-JUL-1994

ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-784-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-104-6

Query Match 20.2%; Score 1233.5; DB 2; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRIRVAERIVATKQGNVTVVVSAMGDTTDLDDLAQQVCPAPPPR 808
DB 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVVCSSAMGDTTDELLELAANVPVPPAR 64

QY 809 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIDVTTCRLQTA 868
DB 65 EMDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNARIIVDTTCRVREA 124

QY 869 LEEGRVVLVAGFGVQSDTKVTLGRGSDTTAVAMAAALGADVCEIYTDVDFGIFADP 928
DB 125 LDEGKICIVAGFGVQNKETRDVTLGRGSDTTAVALLANADVCEIYSDVDGVYADP 184

QY 929 RIVRNARKLDTVTTEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 988
DB 185 RIVPNAQKLEKLSFEEMLELAAGSKILVLRSEYARAFNPLVRSSYSNDPDTLIAGS 244

QY 989 IKDVPMEPDLTGTVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGATDKSEAKVTIVGLISDKGEAAKVFRAADAENIDMVLQNVSSVE 304

QY 1039 HGAERLQGRQD-RHLLHLLP--QTSPPPPKWNTRSETRSASTQLLYDDHIGKVS LI 1095
DB 305 DGTDTITFTCPRADGRRAEMILKKLVQVQ-----NW-----TNVLYDDQVGKVS LV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFNEALRDVNVNIELISTSE 382

RESULT 11
US-08-985-908-5
Sequence 5, Application US/08985908
Patent No. 6004773
GENERAL INFORMATION:
APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHIMURA, AND TSUYOSHI NA
TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,908
FILING DATE: 05-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-325659
FILING DATE: 05-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-985-908-5

Query Match 20.2%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRIRVAERIVATKQGNVTVVVSAMGDTTDLDDLAQQVCPAPPPR 808
DB 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVVCSSAMGDTTDELLELAANVPVPPAR 64

QY 809 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIDVTTCRLQTA 868
DB 65 EMDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNARIIVDTTCRVREA 124

QY 869 LEEGRVVLVAGFGVQSDTKVTLGRGSDTTAVAMAAALGADVCEIYTDVDFGIFADP 928
DB 125 LDEGKICIVAGFGVQNKETRDVTLGRGSDTTAVALLANADVCEIYSDVDGVYADP 184

QY 929 RIVRNARKLDTVTTEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 988
DB 185 RIVPNAQKLEKLSFEEMLELAAGSKILVLRSEYARAFNPLVRSSYSNDPDTLIAGS 244

QY 989 IKDVPMEPDLTGTVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGATDKSEAKVTIVGLISDKGEAAKVFRAADAENIDMVLQNVSSVE 304

QY 1039 HGAERLQGRQD-RHLLHLLP--QTSPPPPKWNTRSETRSASTQLLYDDHIGKVS LI 1095
DB 305 DGTDTITFTCPRADGRRAEMILKKLVQVQ-----NW-----TNVLYDDQVGKVS LV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFNEALRDVNVNIELISTSE 382

RESULT 12

US-08-852-730-14
Sequence 14, Application US/08852730
Patent No. 6090597
GENERAL INFORMATION:
APPLICANT: SEIKO HIRANO, MASAKAZU SUGIMOTO, EIICHI NAKANO,
APPLICANT: MASAKO IZUI, ATSUSHI HAYAKAWA, YASUHIKO YOSHIMURA, AND TSUYOSHI
APPLICANT: NAKAMATSU
TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER AND NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
CITY: ARLINGTON
STATE: VA
ZIP: 22026
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,730
FILING DATE: 05-07-1997

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: JP 8-142812
; FILING DATE: 05-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-852-730-14

Query Match      20.2%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADARRIRVAERIVATKKQGNVVVVVSAMGDTTDDLDDLAQVCPAPPPR 808
DB 5 VQYGGSSLESASERIRVAERIVATKKAGNDVVVVCAMGDTTDELELAANVPVPPAR 64
QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTA 868
DB 65 EMDMLLTAGERISNALVAMAIESLGAESAQFTGSQAGVLTTERHGNARIVDVTPGRVREA 124
QY 869 LEGRVVLVAGFGQVSQDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGIFSADP 928
DB 125 LDEGKICIVAGFGQVKNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184
QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSDRPGTVVWGS 988
DB 185 RIVRNAQKLEKLSFEEMLELAAGVSKTLVRSVEYARAFNPLRVRSYSDNPGTLLIAGS 244
QY 989 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGVATDKSEAKVTVLGIDSKPGEAAKVFRAALADAEINIDMVLQNVSSVE 304
QY 1039 HGAERLQGRQD-RHLLHLP--QTSGPPPKWNTSRSEASQTLLYDDHIGKVSLLI 1095
DB 305 DGTDTITFTCPADGRAMEILKLVQVQ-----TNVLYDDQVGVKSLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 351 GAGMKSHPGVTAFMEALRDVNVNIELISTSE 382

RESULT 13
US-08-985-916-5
; Sequence 5, Application US/08985916
; Patent No. 6221636
; GENERAL INFORMATION:
; APPLICANT: ATSUSHI HAYAKAWA, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; COUNTRY: VA
; ZIP: 22152
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,916
; FILING DATE: 05-DEC-1997
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: JP 8-325658
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-985-916-5

Query Match      20.2%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADARRIRVAERIVATKKQGNVVVVVSAMGDTTDDLDDLAQVCPAPPPR 808
DB 5 VQYGGSSLESASERIRVAERIVATKKAGNDVVVVCAMGDTTDELELAANVPVPPAR 64
QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTA 868
DB 65 EMDMLLTAGERISNALVAMAIESLGAESAQFTGSQAGVLTTERHGNARIVDVTPGRVREA 124
QY 869 LEGRVVLVAGFGQVSQDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFGIFSADP 928
DB 125 LDEGKICIVAGFGQVKNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184
QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSDRPGTVVWGS 988
DB 185 RIVRNAQKLEKLSFEEMLELAAGVSKTLVRSVEYARAFNPLRVRSYSDNPGTLLIAGS 244
QY 989 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGVATDKSEAKVTVLGIDSKPGEAAKVFRAALADAEINIDMVLQNVSSVE 304
QY 1039 HGAERLQGRQD-RHLLHLP--QTSGPPPKWNTSRSEASQTLLYDDHIGKVSLLI 1095
DB 305 DGTDTITFTCPADGRAMEILKLVQVQ-----TNVLYDDQVGVKSLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 351 GAGMKSHPGVTAFMEALRDVNVNIELISTSE 382

RESULT 14
US-08-532-828B-4
; Sequence 4, Application US/0853282B
; Patent No. 5688671
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: OGAWA, Yuri
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: TANAKA, Akiko
; APPLICANT: MATSUI, Hiroshi
; TITLE OF INVENTION: MUTANT ASPARTOKINASE GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS-DOS TEXT EDITOR
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/532,828B
FILING DATE: 27-OCT-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-101450
FILING DATE: 27-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-764-0 PCT
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Corynebacterium glutamicum
STRAIN: AJ3463
US-08-532-828B-4

Query Match 20.1%; Score 1219.5; DB 1; Length 421;
Best Local Similarity 63.8%; Pred. No. 8.3e-101;
Matches 250; Conservative 55; Mismatches 60; Indels 27; Gaps 5;
Qy 749 VQYGGSSVADAEIRIRVAERIVATKKGNDVVVVVVSANGDTTDDLLDLAQOVCPPPPR 808
Db 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVVVCSANGDTTDELLEAAVNPVPPAR 64
Qy 809 ELDMLLTAGERISNALVAVATESLGAHARSPTGSGAGVITGTHGNAKIIDVTGRLQTA 868
Db 65 ENDMLLTAGERISNALVAVATESLGAESAQSFTGSGAGVITGTHGNARIVDVTGVRREA 124
Qy 869 LEEGRVILVAGQGSQDTKVTTLGRGSDTTAVAMAALGADVCEIYTDVDDGIFSDP 928
Db 125 LDEGKICIVAGQGVKTRDTVTTLGRGSDTTAVALLAALNADVCEIYSDVGVYADP 184
Qy 929 RIVRNARKLDTVTESEMLEMAACGAKVLMRCVEYARHNIPVHVRSSYSDEPGTVVGS 988
Db 185 RIVPNAQKLEKUSFEEMLELAAGSKILVLRSEYARAFNPLVRSSYSNDPDTLIAGS 244
Qy 989 IKDVPMDPILTVGAHDSRAKVTIVGLPDPGVAAKVFRAVA-----RRRQHR 1038
Db 245 MEDIPVEEAVLTGATDKSEAKVTVLGSDKPGETAKVFRALADAEINIDMVLQNVSSVE 304
Qy 1039 HGAERLQGRQD-RHLLHLP--QTSPPPKWNTSRSTRSASTQLLYDDHIGKVSLLI 1095
Db 305 DGTDTITFTCPRADGRAMEILKKLVQVG-----TNVLYDDQGVKVSIV 350
Qy 1096 GAGMSEHPGVATFCEALAAVGVNIELISTSE 1127
Db 351 GAGMSEHPGVATFCEALAAVGVNIELISTSE 382

RESULT 15
US-09-252-991A-29720
Sequence 29720, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
CURRENT APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 29720
LENGTH: 461
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29720

Query Match 13.6%; Score 827; DB 4; Length 461;
Best Local Similarity 38.6%; Pred. No. 1.8e-65;
Matches 187; Conservative 86; Mismatches 130; Indels 82; Gaps 7;
Qy 663 ADVRASLERMAPLVDRONAGDVAYRMAPNFDSDIAFLAAQELILSGAQOQNGYTPILH 722
Db 6 APERCAGRAPLDARP-----AVRPASP-----GSRMWOVQKALH 43
Qy 723 RRRREFKARAAEKPAFSDRAGDDAARVQYKVGSSVADAERIRRVARIIVATKKQNDVVV 782
Db 44 GIRR-----FEMALIVQKFGTSGVTVERIEQVAEKVKFREAGDDVVV 87
Qy 783 VVSAMGDTTDDLLDLAQOVCPPPPRELDMLLTAGERISNALVAVATESLGAHARSFTGS 842
Db 88 VVSAMGETNRLTGLANQIMEQVPRELDVVVSTGEQVTIALLSMALIKEGVPAVSYTGN 147
Qy 843 QAGVITGTHGNAKIIDVTGRLQTALEGRVVLVAGFQGVSDTKDVTTLGRGSDTTA 902
Db 148 QVAILTDSATKARILHIDDTHTIRADLKAGRVVVVAGFQGV-DGNGNITTLGRGSDTTG 206
Qy 903 VAMAAALGADVCEIYTDVDDGIFSDADPRIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVE 962
Db 207 VALAAALKADECIYTDVDDGVYTTDPRVVPQARRLDKITFEEMLEMASLGSVKVQIRAVE 266
Qy 963 YARRHNIPVHVRSSYSDEPGTVVVGSIKDVEMEDPILTVGAHDSRAKVTIVGLPDPGY 1022
Db 267 FAGYVPLRVLSFQEGPGLTITIDDESEMEQPIISGTAFNREDAKLTIRGVDPDPGV 326
Qy 1023 AAKVFRAVARRRQ-----HRHGAERLQGRQDRHLLHLPQT 1062
Db 327 AFKILGPISAANVEVDIMVQNVAHNDTDTFTVHRNDVNLAE-----ILKQT 375
Qy 1063 SGPPPKWNTSRSTRSASTQLLYDDHIGKVSLLIGAGVRSHPGVYATFCEALAAVGVNIEL 1122
Db 376 A-----ANIGAREAGTNIKAVSVIGVGNRSHAGVASRNFALAKESINIQM 423
Qy 1123 ISTSE 1127
Db 424 ISTSE 428

Search completed: November 21, 2003, 16:11:41
Job time : 26.7156 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 45.7107 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-52
Perfect score: 6061
Sequence: 1 MQHHHHHTDRVSGNLRIA.....SAATERPRCTGRGRWACQ 1172

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3868	63.8	748	11	US-09-272-975-2
2	3868	63.8	753	11	US-09-272-975-58
3	2289	37.8	739	10	US-09-738-626-6988
4	1615.5	26.7	421	9	US-09-793-306-163
5	1224.5	20.2	421	10	US-09-738-626-3781
6	1224.5	20.2	421	11	US-09-746-660A-24
7	1223.5	20.2	421	15	US-10-226-136-5
8	1217.5	20.1	421	15	US-10-067-974-2
9	1211.5	20.0	421	15	US-10-067-974-16
10	1144.5	18.9	430	15	US-10-156-761-12093
11	770	12.7	156	12	US-10-084-843-227
12	770	12.7	156	12	US-10-193-002-222
13	673	11.1	394	12	US-09-882-227-290
14	377	6.2	564	10	US-09-890-813-17
15	365	6.0	555	10	US-09-890-813-8

16	359	5.9	560	10	US-09-890-813-6	Sequence 6, Appli
17	314.5	5.2	262	12	US-10-301-997-23	Sequence 23, Appl
18	310.5	5.1	449	11	US-09-989-339-23	Sequence 23, Appl
19	306.5	5.1	439	10	US-09-890-813-16	Sequence 16, Appl
20	285.5	4.7	172	15	US-10-226-136-7	Sequence 7, Appli
21	248.5	4.1	281	15	US-09-890-813-12	Sequence 12, Appl
22	241	4.0	440	15	US-10-022-832-34	Sequence 34, Appl
23	197	3.3	4999	10	US-09-976-059-14	Sequence 14, Appl
24	184	3.0	530	15	US-10-156-761-9582	Sequence 9582, Ap
25	179	3.0	304	15	US-10-156-761-14293	Sequence 14293, A
26	173	2.9	5245	12	US-10-329-079-11	Sequence 11, Appl
27	171	2.8	3798	14	US-10-014-717-6	Sequence 6, Appli
28	168	2.8	3970	15	US-10-156-761-10429	Sequence 10429, A
29	166	2.7	3564	15	US-10-156-761-7964	Sequence 7964, Ap
30	161	2.7	6396	12	US-09-940-316B-72	Sequence 72, Appl
31	160	2.6	5245	12	US-10-329-079-45	Sequence 45, Appl
32	151.5	2.5	7746	15	US-10-156-761-7965	Sequence 7965, Ap
33	150	2.5	6146	15	US-10-156-761-10436	Sequence 10436, A
34	149.5	2.5	11877	10	US-09-861-289-6	Sequence 6, Appli
35	149.5	2.5	11877	11	US-09-860-846-6	Sequence 6, Appli
36	149.5	2.5	11877	11	US-09-836-821-6	Sequence 6, Appli
37	149.5	2.5	11877	12	US-10-271-889-49	Sequence 49, Appl
38	149.5	2.5	12199	11	US-09-988-384B-6	Sequence 6, Appli
39	148.5	2.5	4840	15	US-10-156-761-10435	Sequence 10435, A
40	146	2.4	3352	15	US-10-156-761-7961	Sequence 7961, Ap
41	145	2.4	7257	14	US-10-014-717-5	Sequence 5, Appli
42	145.5	2.4	541	15	US-10-156-761-9538	Sequence 9538, Ap
43	145	2.4	1624	16	US-10-080-170-539	Sequence 539, App
44	145	2.4	8360	12	US-10-132-134-34	Sequence 34, Appl
45	144.5	2.4	2472	9	US-09-815-242-5064	Sequence 5064, Ap

ALIGNMENTS

RESULT 1
US-09-272-975-2
; Sequence 2, Application US/09272975
; Publication No. US2003002774A1
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Ronald C.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: TUBERCULOSIS ANTIGENS AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; FILE REFERENCE: 210121.474
; CURRENT APPLICATION NUMBER: US/09/272,975
; CURRENT FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-272-975-2

Query Match	63.8%	Score	3868	DB	11	Length	748
Best Local Similarity	100.0%	Pred. No.	0				
Matches	748	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MQHHHHHTDRVSGNLRARVLYDFVNEALPGTIDPDSFWAGYDKVADLTPOQAL	60				
Db	1	MQHHHHHTDRVSGNLRARVLYDFVNEALPGTIDPDSFWAGYDKVADLTPOQAL	60				
Qy	61	LNARDELQAIQDKWHRRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTTA	120				
Db	61	LNARDELQAIQDKWHRRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTTA	120				
Qy	121	GPOLVVPVFNARFALNAARNGSLYDALYGTDIVPETDGAEGKPTYNKVRGDKVIAYAR	180				
Db	121	GPOLVVPVFNARFALNAARNGSLYDALYGTDIVPETDGAEGKPTYNKVRGDKVIAYAR	180				
Qy	181	KFLDDSVPLSSGSGFGATGTFTVDGOLVVALPDKSTGLANPGOFAGYTGAEPSPTSVLLI	240				

```
Db 181 KFLDDSVPLSSGSGFGDATGFTVODGOLVVALPKSTGLANPGQFAGYTGAAESPTSVLII 240
Qy 241 NHGLHIEIILDPSSQVGTTRAGVKDVIIESAITTIMDFEDSVAADKVLGYRNWLG 300
Db 241 NHGLHIEIILDPSSQVGTTRAGVKDVIIESAITTIMDFEDSVAADKVLGYRNWLG 300
Qy 301 LNKGDIAAADVKGDTAFRLVNRDRNYTAPGGQFTLPGRSLMFRVNVGHLMTNDAIVDT 360
Db 301 LNKGDIAAADVKGDTAFRLVNRDRNYTAPGGQFTLPGRSLMFRVNVGHLMTNDAIVDT 360
Qy 361 DGSEVEGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELF 420
Db 361 DGSEVEGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELF 420
Qy 421 SRVEDVLGLPQNTMKIGIMDEERTTNLKACIKAAADRVPFINTGFLDRTGDEIHTSME 480
Db 421 SRVEDVLGLPQNTMKIGIMDEERTTNLKACIKAAADRVPFINTGFLDRTGDEIHTSME 480
Qy 481 AGPMVRKGTMTKSPWILAYEDHNVDAGLAAGSGRAQVKGKMTMTLMADVMVETKIAQP 540
Db 481 AGPMVRKGTMTKSPWILAYEDHNVDAGLAAGSGRAQVKGKMTMTLMADVMVETKIAQP 540
Qy 541 RAGASTAWVSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 600
Db 541 RAGASTAWVSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 600
Qy 601 IREVDNCCSILGYVVRWDQVGGSKVPDIHDVALMEDRATLRISSOLLANLWLRHGVI 660
Db 601 IREVDNCCSILGYVVRWDQVGGSKVPDIHDVALMEDRATLRISSOLLANLWLRHGVI 660
Qy 661 TSADVRSASLERMAPLVDQRNAGDVAVRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPI 720
Db 661 TSADVRSASLERMAPLVDQRNAGDVAVRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPI 720
Qy 721 LHRRRREFKARAAEKPAAPSDRAGDDAAR 748
Db 721 LHRRRREFKARAAEKPAAPSDRAGDDAAR 748
```

RESULT 2

```
US-09-272-975-58
; Sequence 58, Application US/09272975
; Publication No. US2003002774A1
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Ronald C.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: TUBERCULOSIS ANTIGENS AND METHODS
; FILE REFERENCE: 210121.474
; CURRENT APPLICATION NUMBER: US/09/272,975
; CURRENT FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 753
; TYPE: PR1
; ORGANISM: Mycobacterium tuberculosis
US-09-272-975-58

Query Match 63.8%; Score 3868; DB 11; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 748; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQHHHHHTDRVSGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNAL 60
Db 6 MQHHHHHTDRVSGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNAL 65
Qy 61 LNARDELQAIQDKWHRRRIEPTDMDAYRQFLTEIGYLLPEPDDFTITTSQVDASITTTA 120
Db 66 LNARDELQAIQDKWHRRRIEPTDMDAYRQFLTEIGYLLPEPDDFTITTSQVDASITTTA 125
```

```
Qy 121 GPQLVPLVNARFALNANRWGSLDLYLGTVDVIIPETDGAEGKPTYNKVRGKVIAYAR 180
Db 126 GPQLVPLVNARFALNANRWGSLDLYLGTVDVIIPETDGAEGKPTYNKVRGKVIAYAR 185
Qy 181 KFLDDSVPLSSGSGFGDATGFTVODGOLVVALPKSTGLANPGQFAGYTGAAESPTSVLII 240
Db 186 KFLDDSVPLSSGSGFGDATGFTVODGOLVVALPKSTGLANPGQFAGYTGAAESPTSVLII 245
Qy 241 NHGLHIEIILDPSSQVGTTRAGVKDVIIESAITTIMDFEDSVAADKVLGYRNWLG 300
Db 246 NHGLHIEIILDPSSQVGTTRAGVKDVIIESAITTIMDFEDSVAADKVLGYRNWLG 305
Qy 301 LNKGDIAAADVKGDTAFRLVNRDRNYTAPGGQFTLPGRSLMFRVNVGHLMTNDAIVDT 360
Db 306 LNKGDIAAADVKGDTAFRLVNRDRNYTAPGGQFTLPGRSLMFRVNVGHLMTNDAIVDT 365
Qy 361 DGSEVEGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELF 420
Db 366 DGSEVEGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELF 425
Qy 421 SRVEDVLGLPQNTMKIGIMDEERTTNLKACIKAAADRVPFINTGFLDRTGDEIHTSME 480
Db 426 SRVEDVLGLPQNTMKIGIMDEERTTNLKACIKAAADRVPFINTGFLDRTGDEIHTSME 485
Qy 481 AGPMVRKGTMTKSPWILAYEDHNVDAGLAAGSGRAQVKGKMTMTLMADVMVETKIAQP 540
Db 486 AGPMVRKGTMTKSPWILAYEDHNVDAGLAAGSGRAQVKGKMTMTLMADVMVETKIAQP 545
Qy 541 RAGASTAWVSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 600
Db 546 RAGASTAWVSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 605
Qy 601 IREVDNCCSILGYVVRWDQVGGSKVPDIHDVALMEDRATLRISSOLLANLWLRHGVI 660
Db 606 IREVDNCCSILGYVVRWDQVGGSKVPDIHDVALMEDRATLRISSOLLANLWLRHGVI 665
Qy 661 TSADVRSASLERMAPLVDQRNAGDVAVRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPI 720
Db 666 TSADVRSASLERMAPLVDQRNAGDVAVRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPI 725
Qy 721 LHRRRREFKARAAEKPAAPSDRAGDDAAR 748
Db 726 LHRRRREFKARAAEKPAAPSDRAGDDAAR 753
```

RESULT 3

```
US-09-738-626-6988
; Sequence 6988, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6988
```



```

; LENGTH: 739
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6988

Query Match      37.8%; Score 2289; DB 10; Length 739;
Best Local Similarity 61.3%; Pred. No. 2.2e-188;
Matches 444; Conservative .90; Mismatches 184; Indels 6; Gaps 4;

QY 9 TDRVSGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTTPQNALNARELQ 68
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 20 TERVDAGMVAKVLYDFVTEAVLPVGVDAEKFWSGFAIARDLTFRNELLARRDELQ 79

QY 69 AQIDKWHRRVIEPIDMDAYRQFITEIGYLLPEPDDFTTISGVDAITTTAGQLVVPV 128
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 80 MLIDYHRRN-SGITDQAEYDFLKEIGYLVVEPEAAEIRTONVDTBISSTAGPOLVPI 138

QY 129 LNARFALNAARWGSGLVADLVGTDVTPETDGAEGKPTVYAKVRCGVIAVAREFLDQVSP 188
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 LNARFALNAARWGSGLVADLVGTAIPETDGAEGKPTVYAKVRCGVIAVAREFLDQVSP 198

QY 189 LSSGSFGDATGFTVQDQGVVALPKSTGLANPQCFAGYTGAAESPSTVLLIHNGLHIEI 248
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 LDGASHADVEKYNITDGLAAHIGDSVYRLKNRESYRGFTGNFLDPEAILLETNGLHIEL 258

QY 249 LIDPESQVTTDRAGVKDVIIESAITTIMDPEDSVAADKVLGYRNWGLNKGDLAA 308
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 259 QIDPVHPIGKADKTGLKDIVLESAITTIMDPEDSVAADKVLGYRNWGLNKGDLAA 318

QY 309 AVDKDGTAFRLVNRDRNTAFPGGQFTLPGRSLMFVRNVGHLMTNDIAIVDTDQSEVFEG 368
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 319 EMSXNGRIFTRELANKDRVYIGENGTELVLHGSLLFVRNVGHLMTNDIAIVDTDQSEVFEG 377

QY 369 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKHGPABVAFTELFSRVEDVLG 428
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 378 IMDAVLTTCVCAIPGAPQN---KMRNSRKSGSIYIVKPKHGPABVAFTELFSRVEDVLG 434

QY 429 LPQNTMTKIGWDEERRTTNLCACIKAAADRVVINTGFLDRTGDEHTSMEAGPMVRKG 488
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 435 LPRHLKGVWDEERRTSVNDASIMEVADRLATINTGFLDRTGDEHTSMEAGPMVRKA 494

QY 489 TMSQPTILAYEDHNVDAAGLAFSGRAQVKGKMTMTLMADNVETKIAQPRAGASTAW 548
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 495 DMQAPKQCAVYNNVNDAGIQRLPGKQAIKGNWMTLMMAEMLEKKIQPRAGASTAW 554

QY 549 VPSPTAATHALHYVDVAAVQOGL-AGKRATIEQLLTIPAKELAWAPDEIREVDN 607
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 555 VPSPTGATLHATHVLDVDFVQDELRAAGRRDSLRLNLTPTAPNTNWSSEEEKKEEMDN 614

QY 608 NCQSILGVVVRWVQGVCSKVPDIHDVVALMEDRATLRISSQLLANWLHGVITSADVRA 667
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 615 NCQSILGVVVRWVHGVCSKVPDIHDIDLMEDRATLRISSQLLANWLHGVITSADVRA 674

QY 668 SLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAQELILSGAOPNGYTEPILHRRRE 727
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 675 SLERMAVVDQKAGDEAYRMAPNYDASLAFQAKDLIFEGTKSPSGYTEPILHARRRE 734

QY 728 FKAR 731
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 735 FKAK 738

RESULT 4
US-09-793-306-163
; Sequence 163, Application US/09793306
; Patent No. US20020058200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 163
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: MO-2 (aspartokinase)
US-09-793-306-163

Query Match      26.7%; Score 1615.5; DB 9; Length 421;
Best Local Similarity 85.1%; Pred. No. 1.3e-130;
Matches 338; Conservative 7; Mismatches 35; Indels 17; Gaps 3;

QY 749 VQYGGSSVADAEIRRVAAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCAPPPR 808
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 VQYGGSSVADAEIRRVAAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCAPPPR 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAQVITTTGTHGNAKIIDVTGRLQTA 868
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAQVITTTGTHGNAKIIDVTGRLQTA 124

QY 869 LEEGRVVLVAGQGVSDTKDVTTLGRGGSDTTAVAMAAALGADVCEIITVDVGIIFSADP 928
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 LEEGRVVLVAGQGVSDTKDVTTLGRGGSDTTAVAMAAALGADVCEIITVDVGIIFSADP 184

QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRLCVEYARRHNI PVHVRSSYSDRPGTVVVG 988
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 RIVRNARKLDTVTFEEMLEMAACGAKVLMRLCVEYARRHNI PVHVRSSYSDRPGTVVVG 244

QY 989 IKDVPMDPILTGVAHDRSEAKVTIVGLPDIFGYAAKVFRAVARRRRQRHGAEE----- 1043
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 IKDVPMDPILTGVAHDRSEAKVTIVGLPDIFGYAAKVFRAVARRRRQRHGAEE----- 1043

QY 1044 RLQGRQRQRHHLHLLPQTS---GPPPWKNWTRSETRSASTOLLYDDHIGKVSILGAGMR 1100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 296 VLQNVSKVEDGKTDITFTCSRVDVGPAAVEKLDLSLRNEIGFSQLLYDDHIGKVSILGAGMR 355

QY 1101 SHPGVTATFCEALAAVGVNIELISTSEDQSRCCAAAT 1137
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCDRT 392

RESULT 5
US-09-738-626-3781
; Sequence 3781, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, WASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
```

```
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3781
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3781

Query Match      20.2%; Score 1224.5; DB 10; Length 421;
Best Local Similarity 64.0%; Pred. No. 7.7e-97;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADABRIIRVAERIVATKKQGNVVVVVSAMGDTTDDLDAQQVCPAPPPR 808
DB 5 VQYGGSSLESABRIIRVAERIVATKKAGNDVVVVCAMGDTTDELLELAAAVNPVPPAR 64

QY 809 ELDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNAKIIDVTPGRLOTA 868
DB 65 EMDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNARIVDVTPGRVREA 124

QY 869 LEEGRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFSADP 928
DB 125 LDEGKICIVAGFGVQVNETRDVTTLGRGSGDTTAVAAAAALNADVCEIYSDVDGVYTADP 184

QY 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 988
DB 185 RIVNARKLEKUSFEEMLEMAAGSKILVRSVEYARAFNPVRVSSYNDPFTLIAGS 244

QY 989 IKDVPNEDPILTVGAHDSRSEAKVTIIVGLPDIPGYAAKVFRAVA-----RRRRQHR 1038
DB 245 MEDIPVEEAVLTGVTADTKSEAKVTVLGSDKPGAAKVFRAALADAEINDMVLQNVSSVE 304

QY 1039 HGAERLQGRQD-RHHLHLP--QTSPPFPKWNTRSETRSASTOLLYDDHIGKVSLLI 1095
DB 305 DGTDTITFTCPRSDGRAMEILKLVQVG-----TNVLYDDQVGKVSLLV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 6
US-09-746-660A-24
; Sequence 24, Application US/09746660A
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick
; APPLICANT: Hwang, Byung-Joon
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE OF INVENTION: METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: BGI-121CP2
; CURRENT FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/606740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187970

; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 24
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-746-660A-24

Query Match      20.2%; Score 1224.5; DB 11; Length 421;
Best Local Similarity 64.0%; Pred. No. 7.7e-97;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADABRIIRVAERIVATKKQGNVVVVVSAMGDTTDDLDAQQVCPAPPPR 808
DB 5 VQYGGSSLESABRIIRVAERIVATKKAGNDVVVVCAMGDTTDELLELAAAVNPVPPAR 64

QY 809 ELDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNAKIIDVTPGRLOTA 868
DB 65 EMDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNARIVDVTPGRVREA 124

QY 869 LEEGRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFSADP 928
DB 125 LDEGKICIVAGFGVQVNETRDVTTLGRGSGDTTAVAAAAALNADVCEIYSDVDGVYTADP 184

QY 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 988
DB 185 RIVNARKLEKUSFEEMLEMAAGSKILVRSVEYARAFNPVRVSSYNDPFTLIAGS 244

QY 989 IKDVPNEDPILTVGAHDSRSEAKVTIIVGLPDIPGYAAKVFRAVA-----RRRRQHR 1038
DB 245 MEDIPVEEAVLTGVTADTKSEAKVTVLGSDKPGAAKVFRAALADAEINDMVLQNVSSVE 304

QY 1039 HGAERLQGRQD-RHHLHLP--QTSPPFPKWNTRSETRSASTOLLYDDHIGKVSLLI 1095
DB 305 DGTDTITFTCPRSDGRAMEILKLVQVG-----TNVLYDDQVGKVSLLV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 7
US-10-226-136-5
; Sequence 5, Application US/10226136
; Publication No. US20030054506A1
; GENERAL INFORMATION:
; APPLICANT: OTSUNA, Seiko
; APPLICANT: OGIMOTO, Masakazu
; APPLICANT: IZUI, Masako
; APPLICANT: HAYAKAWA, Atsushi
; APPLICANT: NAKANO, Eiichi
; APPLICANT: KOBAYASHI, Masaki
; APPLICANT: YOSHIMURA, Yasuhiko
; APPLICANT: NAKAMATSU, Tsuyoshi
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/226,136
```

```
;
; FILING DATE: 23-AUG-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,976
; FILING DATE: 8-DEC-1997
; APPLICATION NUMBER: JP 7-140614
; FILING DATE: 07-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-226-136-5

Query Match      20.1%; Score 1217.5; DB 15; Length 421;
Best Local Similarity 63.8%; Pred. No. 3.1e-96;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRRRVAERIVATKKQGNVWVVSAMGDTTDLDDLAQVCPAPPPR 808
DB 5 VQYGGSSLESAERIRNVAERIVATKKAGNDVVVCSAMGDTTDLLELAANVPVPPAR 64
QY 809 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIVDTPGRLOTA 868
DB 65 EMDMLLTAGERISNALVAMAEISLGAESAQFTGSQAGVLTTERHGNARIVDTPGRVREA 124
QY 869 LEEGRVVLVAGFGVQSQDTKVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFPSADP 928
DB 125 LDEGKICIVAGFGVQVKNKTRDVTTLGRGSGDTTAVALAAALNADVCEIYSDVDGYTADP 184
QY 929 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 988
DB 185 RIVPNAQKLEKLSFEEMLELAAGSKILVRSVEYARAFNVPLRVRSYSNDPGTLLAGS 244
QY 989 IKDVPMDPILTVGAHDSERAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGATDKSEAKVTVLGSDKPGEAAKVFRALADAEINIDMVLQNVSSVE 304
QY 1039 HGAERLQGRGQD--RHHLLLP--QTSQPPPKWTRSETRSASTQLLYDDHIGKVSLLI 1095
DB 305 DGTDTITTCPRADGRAMEILKQLVQG-----TNVLYDDQVGKVSLLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 9
US-10-067-974-16
; Sequence 16, Application US/10067974
; Publication No. US20030055232A1
; GENERAL INFORMATION:
; APPLICANT: Li, Lhing-Yew
; APPLICANT: Trei, Kelli J.
; TITLE OF INVENTION: Polynucleotide Constructs for Increased Lysine Production
; FILE REFERENCE: 1533.2640001
; CURRENT APPLICATION NUMBER: US/10/067,974
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,183
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 16
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus Sequence of Protein Sequence Alignment
; NAME/KEY: MISC FEATURE
; LOCATION: (40)..(40)
; OTHER INFORMATION: May be either Cys or Val
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (317)..(317)
; OTHER INFORMATION: May be either Ser or Ala
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (345)..(345)
; OTHER INFORMATION: May be either Gly or Asp
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (380)..(380)
; OTHER INFORMATION: May be either Thr or Ile
;

;
; FILING DATE: 23-AUG-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,976
; FILING DATE: 8-DEC-1997
; APPLICATION NUMBER: JP 7-140614
; FILING DATE: 07-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-226-136-5

Query Match      20.2%; Score 1223.5; DB 15; Length 421;
Best Local Similarity 64.0%; Pred. No. 9.4e-97;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRRRVAERIVATKKQGNVWVVSAMGDTTDLDDLAQVCPAPPPR 808
DB 5 VQYGGSSLESAERIRNVAERIVATKKAGNDVVVCSAMGDTTDLLELAANVPVPPAR 64
QY 809 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIVDTPGRLOTA 868
DB 65 EMDMLLTAGERISNALVAMAEISLGAESAQFTGSQAGVLTTERHGNARIVDTPGRVREA 124
QY 869 LEEGRVVLVAGFGVQSQDTKVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFPSADP 928
DB 125 LDEGKICIVAGFGVQVKNKTRDVTTLGRGSGDTTAVALAAALNADVCEIYSDVDGYTADP 184
QY 929 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 988
DB 185 RIVPNAQKLEKLSFEEMLELAAGSKILVRSVEYARAFNVPLRVRSYSNDPGTLLAGS 244
QY 989 IKDVPMDPILTVGAHDSERAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGATDKSEAKVTVLGSDKPGEAAKVFRALADAEINIDMVLQNVSSVE 304
QY 1039 HGAERLQGRGQD--RHHLLLP--QTSQPPPKWTRSETRSASTQLLYDDHIGKVSLLI 1095
DB 305 DGTDTITTCPRADGRAMEILKQLVQG-----TNVLYDDQVGKVSLLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 8
US-10-067-974-2
; Sequence 2, Application US/10067974
; Publication No. US20030055232A1
; GENERAL INFORMATION:
; APPLICANT: Li, Lhing-Yew
; APPLICANT: Trei, Kelli J.
; TITLE OF INVENTION: Polynucleotide Constructs for Increased Lysine Production
; FILE REFERENCE: 1533.2640001
; CURRENT APPLICATION NUMBER: US/10/067,974
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,183
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 421
; TYPE: PRT
;
```

US-10-067-974-16

Query Match 20.0%; Score 1211.5; DB 15; Length 421;
Best Local Similarity 63.5%; Pred. No. 1e-95;
Matches 249; Conservative 55; Mismatches 61; Indels 27; Gaps 5;

QY 749 VQYGGSSVADABIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLDAQVCPAPPPR 808
DB 5 VQYGGSSLESABIRIRVAERIVATKKAGNVNVVVSAMGDTTDELLELAANVPVPPAR 64

QY 809 ELDMLTAGERISNALVANAISLGAHARSFTGSQAGVITTGTHGNAKIIVDTPGRLOTA 868
DB 65 EMDMLTAGERISNALVANAISLGAHARSFTGSQAGVITTEHGNARIVDTPGRVREA 124

QY 869 LEGRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVVGIFSDAP 928
DB 125 LDEGKICIVAGFGVQVNETRDVTTLGRGSGDTTAVAMAAALNADVCEIYSDVDGVYTADP 184

QY 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 988
DB 185 RIVNARKLEKTSFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 244

QY 989 IKDVPMPILTGVAHDSRSEAKVTIYGLPDIPGYAAKVFRAVA--RRRQHR 1038
DB 245 MEDIPVEEALVTGATDKSEAKVTYVLGSDKPGCEAAKVFRAADAEINIDVWLQNVSSVE 304

QY 1039 HGAERLQGRGROD-RHLLHLLP--QTSGPPPKWTRSETRSASTOLLYDDHIGKVSIL 1095
DB 305 DGTDTITFTCPRXDGRAMEILKQLVQVQ-----TNVLYDDQVXKVSILV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 351 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 382

RESULT 10

US-10-156-761-12093
; Sequence 12093, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12093
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12093

Query Match 18.9%; Score 1144.5; DB 15; Length 430;
Best Local Similarity 60.2%; Pred. No. 6.5e-90;
Matches 239; Conservative 54; Mismatches 75; Indels 29; Gaps 4;

QY 749 VQYGGSSVADABIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLDAQVCPAPPPR 808
DB 5 VQYGGSSVADABIRIRVAERIVATKKAGNVNVVVSAMGDTTDELLELAANVPVPPAR 64

QY 809 ELDMLTAGERISNALVANAISLGAHARSFTGSQAGVITTGTHGNAKIIVDTPGRLOTA 868
DB 65 EFDMLTAGERISNALVANAISLGAHARSFTGSQAGVITTSVHNAKIIVDTPGRIRTA 124

QY 869 LEGRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVVGIFSDAP 928
DB 125 LDEGKICIVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALDAEVCEIYTDVVGIFSDAP 184

QY 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 988
DB 185 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 241

QY 989 IKDVPMPILTGVAHDSRSEAKVTIYGLPDIPGYAAKVFRAVA--RRRQHR 1034
DB 242 -SNTFLVQKQAGQEQVEQAIIISGVVADTSEAKVTYVGVDPKPGCEAAASIFRAIDAENV 300

QY 1035 ----RQHRHGAERLQGRGROD-RHLLHLLP--QTSGPPPKWTRSETRSASTOLLYDDHIG 1090
DB 301 IDMVVQNVSAASTGLTDIS-----FTLPEKTEGRKAIDALEKAKSVIGFDSLYDDQIG 353

QY 1091 KVSILGAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 354 KISLVGAGMRSHPGVTATFCEALAAVGVNIELISTSE 390

RESULT 11

US-10-084-843-227
; Sequence 227, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 227:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 227:

US-10-084-843-227

Query Match 12.7%; Score 770; DB 12; Length 156;
 Best Local Similarity 100.0%; Pred. No. 2.7e-58;
 Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 750 QYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDDLAQVCPAPPPRE 809
 Db 1 QYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDDLAQVCPAPPPRE 60
 QY 810 LDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPGRLOTAL 869
 Db 61 LDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPGRLOTAL 120
 QY 870 BEGRVVLVAGFGQVSQDTKDVTTTLGRGSGDTTAVAM 905
 Db 121 BEGRVVLVAGFGQVSQDTKDVTTTLGRGSGDTTAVAM 156

RESULT 12

US-10-193-002-222
 ; Sequence 222, Application US/10193002
 ; Publication No. US20030135026A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

Skeiky, Yasir A.W.
 Dillon, Davin C.
 Campos-Neto, Antonia
 Houghton, Raymond
 Vedvick, Thomas S.
 Twardzik, Daniel R.
 Lodes, Michael J.
 Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

TUBERCULOSIS

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESS: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 City: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/193,002
 FILING DATE: 10-Jul-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 222:

SEQUENCE CHARACTERISTICS:

LENGTH: 156 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 222:

US-10-193-002-222

Query Match 12.7%; Score 770; DB 12; Length 156;

Best Local Similarity 100.0%; Pred. No. 2.7e-58;
 Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 QYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDDLAQVCPAPPPRE 809
 Db 1 QYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDDLAQVCPAPPPRE 60
 QY 810 LDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPGRLOTAL 869
 Db 61 LDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPGRLOTAL 120
 QY 870 BEGRVVLVAGFGQVSQDTKDVTTTLGRGSGDTTAVAM 905
 Db 121 BEGRVVLVAGFGQVSQDTKDVTTTLGRGSGDTTAVAM 156

RESULT 13

US-09-882-227-290

; Sequence 290, Application US/09882227

; Publication No. US20030159396A1

GENERAL INFORMATION:

APPLICANT: Kleanthous, Harold

APPLICANT: Al-Garawi, Amal

APPLICANT: Miller, Charles

APPLICANT: Tomb, Jean-Francois

APPLICANT: Coomen, Raymond P.

TITLE OF INVENTION: Identification of Polynucleotides

TITLE OF INVENTION: Encoding No. US20030159396A1

TITLE OF INVENTION: Genome

FILE REFERENCE: 06132/047002

CURRENT APPLICATION NUMBER: US/09/882,227

CURRENT FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 08/902,615

PRIOR FILING DATE: 1997-07-29

NUMBER OF SEQ ID NOS: 638

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 290

LENGTH: 394

TYPE: PRT

ORGANISM: Helicobacter pylori

US-09-882-227-290

Query Match 11.1%; Score 673; DB 12; Length 394;

Best Local Similarity 40.1%; Pred. No. 3e-49;

Matches 151; Conservative 83; Mismatches 119; Indels 24; Gaps 6;

QY 761 ERIRVAERIVATKKQNDVVVVVSAMGDTTDDLDDLAQVCPAPPPRELDMLLTAGERI 820
 Db 5 ERHNVARVLESVTLLGHQVWVVVSAMSGTDLLEFGKNFHNPNKREMDRIVSVGELY 64

QY 821 SNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPGRLOTALBEGRVVLVAGF 880
 Db 65 SSAALSVALERYGHRALSLSKEAGILTSSEFQNAVIOSTDKRITELLEKNYIVVIAGF 124

QY 881 QGVSDTKDVTTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADPRIVRNARKLDTV 940
 Db 125 QG-ADIOGETTTLGRGSGDLSAVALAGAKLAHLCEIYTDVGVYTTDFRIEERAKIAQI 183

QY 941 TFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVSSYSDRPGTWWVG--SIKDVPMEDPI 998
 Db 184 SYDEMELASGAKVLLNRSVLAKLSVKLVTRNSFNHSEGLIIVAEKDFKGERMETPI 243

QY 999 LTGVADHRSEAKYTIIVGLPIPGYAAKVFRAVARRRRQHRGAERLQGRGRDRHLHL 1058
 Db 244 VSGIALDKNQARVSMEGVEDRPGIAAEIFGALA---EYRINVDIMIVQTTGRDKTDLDF 299

QY 1059 -LPQTSPPPKWNRSTRSASTOLL-----YDDHIGKVSILGAGMESHQVATATFC 1110
 Db 300 TIVKTQ-----IETQALKPFLAQMDSIDYDENTAKVISIVGVGMKSHSGVASIAF 350

QY 1111 EALAAVGVNIELSTSE 1127

351 KALAKDNINIMMISTSE 367

RESULT 14

US-09-890-813-17
 ; Sequence 17, Application US/09890813
 ; Publication No. US20020183486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: E.I. du Pont de Nemours and Company
 ; TITLE OF INVENTION: Aspartate Kinase
 ; FILE REFERENCE: BB1430 PCT
 ; CURRENT APPLICATION NUMBER: US/09/890,813
 ; CURRENT FILING DATE: 2001-08-02
 ; PRIOR APPLICATION NUMBER: 60/172944
 ; PRIOR FILING DATE: 1999-12-21
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 17
 ; LENGTH: 564
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 US-09-890-813-17

Query Match 6.2%; Score 377; DB 10; Length 564;

Best Local Similarity 27.0%; Pred. No. 2e-23;
 Matches 128; Conservative 71; Mismatches 165; Indels 110; Gaps 13;

```

QY 749 VKYGGSSVADAEIRIRVAERIVATKKQGNVTVVVSAMGDTTDDL----- 795
DB 87 VMKFGSSVASADRMKEVATLILSPFEE--RPVIVLSANGKTKNLLAGEKAVSCGVIN 144
QY 796 -----DLAQOVCPAPPREL-----DMLTAG 817
DB 145 VSSIEELCFIKDLHLRTVDQGVDSVISKHLEELBOLLKGIAMMKELTKRTQDYLVSGF 204
QY 818 ERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVT---PGRLO-TALEEG 872
DB 205 ECMSTRIFAAYLNKIGVKARQYDAFEGITTTDFNADILEATYPVAKRHLHGWLSDP 264
QY 873 RVVLVAGFGVSGDTKDVTLGRGSDTTAVAAALGADVCEIYTDVGIFSDPRIVR 932
DB 265 AIAIVTGLGKARKSKAVITTLGRGSDLTATTIGKALGPEIQVWMDVGLTCDPNIVP 324
QY 933 NARKLTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSR--PGTVVVSIGD 991
DB 325 KAPVPYLTDEAAELAYFGAQLHFPQSMRPARESDIPVRVKSNSYNPKAPGLIT---KA 381
QY 992 VPMEDEPILTVGAHDSRSEAKVTIVGLPDIP--GYAAKVFRVAVARRRRQRHGHGAERLQGRG 1049
DB 382 RDMSKAVLTSIVLKRNVMTLDIASTRMLQGYGLAKVFSIFEELGTSVDVATSEVSVL 441
QY 1050 RQRHHLHLLPQTSGPPKWNTRSTRSASTQLLYDDHIGK-----VS 1093
DB 442 TLD-----PSKLMRELIOQASEL-----DHWVEELEKIAVNVNLLQNRSTIS 483
QY 1094 LIGAGMRSHPGVATPCEALAAVGNVIELIS-----TSEDQSRCCAA 1136
DB 484 LIGNVQRSSU-ILERSLVRLTIGTVQMISQASKVNLISLVVNDSEAEQCVRA 536

```

RESULT 15

US-09-890-813-8
 ; Sequence 8, Application US/09890813
 ; Publication No. US20020183486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: E.I. du Pont de Nemours and Company
 ; TITLE OF INVENTION: Aspartate Kinase
 ; FILE REFERENCE: BB1430 PCT
 ; CURRENT APPLICATION NUMBER: US/09/890,813
 ; CURRENT FILING DATE: 2001-08-02
 ; PRIOR APPLICATION NUMBER: 60/172944
 ; PRIOR FILING DATE: 1999-12-21
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: Microsoft Office 97

; SEQ ID NO 8
 ; LENGTH: 555
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 US-09-890-813-8

Query Match 6.0%; Score 365; DB 10; Length 555;
 Best Local Similarity 29.9%; Pred. No. 2.1e-22;
 Matches 103; Conservative 58; Mismatches 119; Indels 64; Gaps 7;

```

QY 743 GDDAARYQKYGGSSVADAEIRIRVAERIVATKKQGNVTVVVSAMGDTTD----- 752
DB 80 GDQLSVVMKFGSSVSSAARMAEVAGLILTFPEE--RPVVVLSAMGKTTNNLLAGEKAV 137
QY 793 -----DLDLAQOVCPA-----PPPRELD 811
DB 138 GCGVIHVSEIEEEMNMVKSLLHKTVDLGLPRSVIQDMLDELEQLLKGIAVMKELTPRTSD 197
QY 812 MLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTTPGSLQATL-- 869
DB 198 YLVSFGECSMTRIPSAVLNKRIVKARQYDAFDIGFITDDEFGNADILEATYPVAKRELHG 257
QY 870 ---BEGRVVLVAGFGVSGDTKDVTLGRGSDTTAVAAALGADVCEIYTDVGIFSA 926
DB 258 DWIQDPAIPVVVTGFLGKWKSGAVTTLGRGSDLTATTIGKALGRLREIQWMDVGLTLC 317
QY 927 DPRIVRVARKLDVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSR--PGTVV 985
DB 318 DPNIYPHAKTVPYLTTFEATELAYFGAQLHFPQSMRPAREGDIPVRVKSNSYNPKAPGLTI 377
QY 986 VGSIKDVPMEDPILTVGAHDSRSEAKVTIVGLPDIP--GYAAKV 1027
DB 378 T---RQRDMDKVLTSTVLKSNVTMLDIVSTRMLQGYGLARVF 418

```

Search completed: November 21, 2003, 16:38:14
 Job time : 46.7107 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 24.4499 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-52
Perfect score: 6061
Sequence: 1 MQHHHHHTDRVSVGNLRIA.....SAATRRPCTGRGRWACQ 1172
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3810	62.9	741	2 F70722	probable glcB prot
2	3117.5	51.4	731	2 T44752	probable malate sy
3	2569	42.4	725	2 H83586	malate synthase G
4	2380	39.3	744	2 AH2582	malate synthase G
5	2380	39.3	744	2 F97364	malate synthase G
6	2318	38.2	728	2 AE3299	malate synthase (E
7	2289	37.8	739	2 I40715	malate synthase (E
8	2258	37.3	727	2 E83916	malate synthase BH
9	2154.5	35.5	723	2 S51788	malate synthase (E
10	1615.5	26.7	421	2 F70794	probable ask prote
11	1545.5	25.5	421	2 G87199	aspartokinase (imp
12	1478.5	24.4	421	2 S42422	aspartate kinase (
13	1279.5	20.3	421	2 S15276	aspartate kinase (
14	1141	18.8	425	2 T33383	probable aspartoki
15	823	13.6	412	2 C83531	aspartate kinase a
16	816.5	13.5	600	2 S76764	hypothetical prote
17	785	13.0	411	2 A48946	aspartate kinase (
18	778	12.8	407	2 JC4640	aspartate kinase (
19	776.5	12.8	606	2 AE2261	aspartate kinase (
20	761	12.6	405	2 H81865	aspartate kinase (
21	760	12.5	405	2 F81076	aspartokinase, alp
22	757.5	12.5	412	2 H84036	aspartokinase II a
23	742.5	12.3	400	2 E81405	aspartate kinase (
24	736.5	12.2	739	2 H72364	aspartokinase II -
25	734.5	12.1	415	2 B70399	aspartokinase - Aq
26	724.5	12.0	401	2 G72245	aspartokinase II -
27	721	11.9	405	2 H71843	aspartokinase 2 al
28	712.5	11.8	408	2 A29314	aspartate kinase (
29	711	11.7	405	2 E64673	aspartokinase - He

RESULT 1
F70722
Probable glcB protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70722
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A: Title: Deciphering the Biology of Mycobacterium tuberculosis from the complete genome: A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70722
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-741 <COL>
A: Cross-references: GB:278020; GB:AL123456; NID: g3261625; PIDN: CAB01465.1; PID: e257679;
A: Experimental source: strain H37RV
C: Genomics:
A: Gene: glcB

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
62.9%	100.0%	740	0	0	0	0
Qy	9	TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFVAGVQKVVADLTPOQALLNARDELQ	68			
Db	2	TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFVAGVQKVVADLTPOQALLNARDELQ	61			
Qy	69	AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPPDDFTITTSVDABITTTAGPOLVVPV	128			
Db	62	AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPPDDFTITTSVDABITTTAGPOLVVPV	121			
Qy	129	LNARFALNANARWGLYDALYGTDPETDGAEGKPTYNKVRGDKVIAYARKFLDDSV	188			
Db	122	LNARFALNANARWGLYDALYGTDPETDGAEGKPTYNKVRGDKVIAYARKFLDDSV	181			
Qy	189	LSSGSFGDAGFTVQDQQLVVALPKSTGLANPQFAGYTGAAESPTSVLLINHLHIEI	248			
Db	182	LSSGSFGDAGFTVQDQQLVVALPKSTGLANPQFAGYTGAAESPTSVLLINHLHIEI	241			
Qy	249	LIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAAVDAADKVLGYRNWGLNKGDLAA	308			
Db	242	LIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAAVDAADKVLGYRNWGLNKGDLAA	301			
Qy	309	AVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRLSMFVRNVGHMLTNDALVDTDGSSEVFEG	368			
Db	302	AVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRLSMFVRNVGHMLTNDALVDTDGSSEVFEG	361			
Qy	369	IMDALFTGLTAHCLKASDVNGPLINSRTGSIIVVKKMGHPAEVATCFLFSRVEDVLG	428			

Db 362 IMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKHGPAAVAFTELSRVEDVLG 421
 QY 429 LPQNTMKIGIMDEERTTNLTKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPVVRG 488
 Db 422 LPQNTMKIGIMDEERTTNLTKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPVVRG 481
 QY 489 TMRKSPQWILAYEDHNVDAGLAAGFSRAQVKGKGMWMTLMADMVETKIAQPRAGASTAW 548
 Db 482 TMRKSPQWILAYEDHNVDAGLAAGFSRAQVKGKGMWMTLMADMVETKIAQPRAGASTAW 541
 QY 549 VPSTPAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPEIREVDNN 608
 Db 542 VPSTPAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPEIREVDNN 601
 QY 609 CQSTILGVVVRWVDQVGCCKVPDIHDVVALMEDRATIRISSQLLANWLHGVITSADVRAS 668
 Db 602 CQSTILGVVVRWVDQVGCCKVPDIHDVVALMEDRATIRISSQLLANWLHGVITSADVRAS 661
 QY 669 LERMAPLVDRQNGADVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPILHRRREF 728
 Db 662 LERMAPLVDRQNGADVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPILHRRREF 721
 QY 729 KARAEEKPAPSDRAGDDAAR 748
 Db 722 KARAEEKPAPSDRAGDDAAR 741

RESULT 2
 T44752
 probable malate synthase (EC 4.1.3.2) G [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 05-May-2000
 C:Accession: T44752
 R:Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, October 1997
 A:Reference number: Z22833
 A:Accession: T44752
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-731 <FAR>
 A:Cross-references: EMBL:AL008609; PIDN:CAA15459.1
 A:Experimental source: cosmid B1788
 C:Genetics:
 A:Note: glcB
 C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 51.4%; Score 3117.5; DB 2; Length 731;
 Best Local Similarity 82.0%; Pred. No. 2.7e-179;
 Matches 597; Conservative 54; Mismatches 72; Indels 5; Gaps 2;

QY 9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFAGVGVKQVADLTPOQALLNARDELQ 68
 Db 2 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFAGVGVKQVADLTPOQALLNARDELQ 61
 QY 69 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVWPV 128
 Db 62 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVWPV 121
 QY 129 LNARFALNANARWGSILYDALYGTDTVPETDGAEGKPTYNKVGDKVIAYARKEFLDDSV 188
 Db 122 LNARFALNANARWGSILYDALYGTDTVPETDGAEGKPTYNKVGDKVIAYARKEFLDDSV 181
 QY 189 LSSGSFGDATGFTVQDQQLVVALPDKSTGLANPGOPAGYTGAAESPT-SVLLINHLGHE 247
 Db 182 LASDSWTNATGVSIFDQQLQIAGTSTGLASPEKPVGNRQLRSSNWSVLLANHLGHE 241
 QY 248 ILIDPESQVGTTRDRAGVKDVILESATITIMDFEDSVAADADKVLGRNWLGLNKGLDIA 307
 Db 242 VLIDPESQVGTTRDRAGVKDVILESATITIMDFEDSVAADADKVLGRNWLGLNKGLDIA 301
 QY 308 AAVDKDGTAFRLVNRDRNRYTPGGQQTLPGRSLMFVRNVGHMTNDAIV----DTGGS 363
 Db 302 EEWKDKGKFTFRLVNRDRNRYTPGGQQTLPGRSLMFVRNVGHMTNDAIV----DTGGS 361

QY 364 EVFSGIMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKHGPAAVAFTELSR 423
 Db 362 EVFSGIMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKHGPAAVAFTELSR 421
 QY 424 EDVLGLPQNTMKIGIMDEERTTNLTKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGP 483
 Db 422 EDVLGLPQNTMKIGIMDEERTTNLTKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGP 481
 QY 484 MYRGTMTKSPQWILAYEDHNVDAGLAAGFSRAQVKGKGMWMTLMADMVETKIAQPRAG 543
 Db 482 MIRKGMKNSWIKAYEDANDVIGLAAAGFKGAQIGKGMWMTLMADMVETKIAQPRAG 541
 QY 544 ASTAWVSPSTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPEIRE 603
 Db 542 ASTAWVSPSTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPEIRE 601
 QY 604 EVDNNCQSIILGVVVRWVDQVGCCKVPDIHDVVALMEDRATIRISSQLLANWLHGVITSA 663
 Db 602 EVDNDCQSIILGVVVRWVDQVGCCKVPDIHDVVALMEDRATIRISSQLLANWLHGVITSE 661
 QY 664 DVRSASLERMAPLVDRQNGADVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPILHR 723
 Db 662 DVRSASLERMAPLVDRQNGADVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPILHR 721
 QY 724 RRREFKAR 731
 Db 722 RRREFKAR 729

RESULT 3
 H83586
 malate synthase G PA0482 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: H83586
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Bri-
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: H83586
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-725 <STC>
 A:Cross-references: GB:AF004485; GB:AE004091; NID:g99946332; PIDN:AAG03871.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: glcB; PA0482

Query Match 42.4%; Score 2569; DB 2; Length 725;
 Best Local Similarity 69.1%; Pred. No. 2.3e-146;
 Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;

QY 9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFAGVGVKQVADLTPOQALLNARDELQ 68
 Db 2 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFAGVGVKQVADLTPOQALLNARDELQ 61
 QY 69 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVWPV 128
 Db 62 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVWPV 121
 QY 129 LNARFALNANARWGSILYDALYGTDTVPETDGAEGKPTYNKVGDKVIAYARKEFLDDSV 188
 Db 122 LNARFALNANARWGSILYDALYGTDTVPETDGAEGKPTYNKVGDKVIAYARKEFLDDSV 181
 QY 189 LSSGSFGDATGFTVQDQQLVVALPDKS-TGLANPGOPAGYTGAAESPTSVLLINHLGHE 247
 Db 182 LESGSHVDATSYSVKNGALVVALKNGSETGLKQAGFLAFOGDAAPQVALLKHGLHFE 241
 QY 248 ILIDPESQVGTTRDRAGVKDVILESATITIMDFEDSVAADADKVLGRNWLGLNKGLDIA 307

Db 242 IQIDPSSVPGQDAAAGVKVGLMEALTTIMDCEDSVAAVADADDKVIYVNWGLMKGDLA 301
QY 308 AAVDXDGTAFRLVRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMNDIAIVDTGSEVFE 367
Db 302 EVSFGGSGTFRITWNPDRVYTRADGSELTLHGRSLFVRNVGHLMNDIAILDGCGNEVPE 361
QY 368 GIMDALFTGLIAIHGLKASDVNGPLI--NSRTGSIYIVKPKMHGPAEVAFTCELFSEVED 425
Db 362 GJQDGLFTSLIAIK-----DLNGNTRKNSRTGSIYIVKPKMHGPEEAFTNEIFGRVED 416
QY 426 VLGLPQNTWKIGIMDEERTTNVNLKACIKAAADRVPVINTGFLDRTGDEIHTSMEAGPMV 485
Db 417 VLGLPRLNTLKVIGIMDEERTTNVNLKACIKAAADRVPVINTGFLDRTGDEIHTSMEAGAVV 476
QY 486 RGTWKSQPTILAYEDHNVDAAGLAAGFSGRAQVKGKMTMELMADVETKIAOPRAGAS 545
Db 477 RKGAKSKKWIYAYENNNDVGLATGLOKQAGIIGKMWAMPDLMAAMLEQKIGHPLAGAN 536
QY 546 TAWPSPATAATHALHYHQVDVAAVQOGLAGKRRTIEQLLTIPLAKELAWAPDEIRBEV 605
Db 537 TAWPSPATAATHALHYHKVDVFAQAELAKRTPASVDDILTIPLAPNTNTASEIKNEV 596
QY 606 DNNCSILGIVVRVWDQGVGSKVPDIHDVAMDEDRATLRISQQLANLWLRHGVITSADV 665
Db 597 DNNAGILGIVVRVWDQGVGSKVPDIHDVAMDEDRATLRISQQLANLWLRHGVISQBV 656
QY 666 RASLERMAPLVRQNAVGVAYRPMAPNFDSDIAFLAAOELILSGAQONGYTEPILHRRR 725
Db 657 VESLXMAVVRQNAVGSYRPMAPNFDSDVAFQAELVVEGTRQPNGYTEPVHLHRR 716
QY 726 REFKA 731
Db 717 REFKA 722

RESULT 4
AH2582
malate synthase G [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH2582
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan; Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH2582
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41078.1; PID:gi7738367; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: glCB
A:Map position: circular chromosome

Query Match 39.3%; Score 2380; DB 2; Length 744;
Best Local Similarity 63.5%; Pred. No. 5.3e-135;
Matches 461; Conservative 91; Mismatches 170; Indels 4; Gaps 2;

QY 8 HTRDVSUGNLRARVLYDFVNEALPGTDIDPDPSFWAGVDKVVADLTQNOALLNARDEL 67
Db 22 HVSRTDKFGLSIDRLYAFLTDEVLPOTGLDSTFFEGFSIAVHLSPKNRELLAKRDAL 81
QY 68 QAOIKDWHRRVRIEPIEDMDAYROFLTEIGYLLPEPDDFTTTTSGVDIAITTTAGPOLVVP 127
Db 82 QKIDGWRENG-APSDFDAYEAFKEIGYLLPEGPGFKEVNNVDPPIAVVAGPOLVVP 140
QY 128 VLNAPALMAANARWGSLYDALYGTDTVITPDTGAEGKPTYNKVRGDKVIAYARFLLDSV 187

Db 141 VMNARVALNAANARWGSLYDALYGTDAISDADGAEEKRGYNPKRGDKVIAWENFLDESA 200
QY 188 PLSGSGFGDGTATVQDQGLVVALPKSTGLANPGQFAGYTGAESPTSVLLINHGHLIE 247
Db 201 PLETGWSVDVTGNIADGLLQLAIGAATTLGLKDAVQFKGFSGAAPATILLKNGHLIHE 260
QY 248 ILDPSSQVQTTDRAGVKDVIIESALTTIMDRSDSVAAVDAADKVLGYRWGLNGLKGDLA 307
Db 261 IVIDPSTEIKSPRAGISDVILESALTTIMDCEDSVAAVDAEDKVLVYGNWGLMGRDILT 320
QY 308 AAVDKOQTAFRLVRNDRNYTAPGGQFTLPGRSLLMFVRNVGHLMNDIAIVDTGSEVFE 367
Db 321 EAVSKGNTFTRLNPDYVYTPDGSALTLPGRSLLMFVRNVGHLMNPAILDRDGDVPE 380
QY 368 GIMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELFSEVEDVL 427
Db 381 GIMDAVVTALIAIYDVGPS---GRRONSRAGSYVVRKMHGPEEFAFANEIIFARVENLV 437
QY 428 GLPQNTWKIGIMDEERTTNVNLKACIKAAADRVPVINTGFLDRTGDEIHTSMEAGPMVK 487
Db 438 GMAPNTWKIGIMDEERTTNVNLKESIRAAKDRVPVINTGFLDRTGDEIHTSMEAGPMVK 497
QY 488 GTWKSQPTILAYEDHNVDAAGLAAGFSGRAQVKGKMTMELMADVETKIAOPRAGASTA 547
Db 498 GDMKQAAWIAAYENWNVDIGLECLSGHAQIGKMWAMPDLMAAMLEQKIAHPKAGANTA 557
QY 548 WPSPTAATLHALHYHQVDVAAVQOGLAGKRRTIEQLLTIPLAKELAWAPDEIRSEVON 607
Db 558 WPSPTAATLHALHYHKVDVAAVQOGLKSRGAKLSILSVPAVRPNWTPEIQRELDN 617
QY 608 NCOSILGIVVRVWDQGVGSKVPDIHDVAMDEDRATLRISQQLANLWLRHGVITSADVRA 667
Db 618 NAQOILGIVVRVWDQGVGSKVPDIHDVAMDEDRATLRISQOAHMANLWLRHGVVTEAIIK 677
QY 668 SLERMAPLVRQNAVGVAYRPMAPNFDSDIAFLAAOELILSGAQONGYTEPILHRRRRE 727
Db 678 TMKRMAAVDTQNAVDPAYLPMSADFGSVAFQAELVVLKGREQNGYTEPVLHRRLE 737
QY 728 FKARAA 733
Db 738 LKAKQA 743

RESULT 5
F97364
malate synthase G (PA0482) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: F97364
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: F97364
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK85871.1; PID:gl5154912; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C 78
A:Map position: circular chromosome

Query Match 39.3%; Score 2380; DB 2; Length 744;
Best Local Similarity 63.5%; Pred. No. 5.3e-135;
Matches 461; Conservative 91; Mismatches 170; Indels 4; Gaps 2;

QY 8 HTRDVSUGNLRARVLYDFVNEALPGTDIDPDPSFWAGVDKVVADLTQNOALLNARDEL 67
Db 22 HVSRTDKFGLSIDRLYAFLTDEVLPOTGLDSTFFEGFSIAVHLSPKNRELLAKRDAL 81
QY 68 QAOIKDWHRRVRIEPIEDMDAYROFLTEIGYLLPEPDDFTTTTSGVDIAITTTAGPOLVVP 127

Db 82 QEKIDGWYRENG-APSDFDAYEAFLEKIGYLLPEGPGFKVETNNVDPEIAVAVAGPQLVVP 140
QY 128 VLNARFALNANRWGSLYDALYGTQVI PETDGAERKPTYNKVRGDKVIAYARKFLDDSV 187
Db 141 VMARYALNANRWGSLYDALYGTDAISDADGAERKGRYPNKRGDVKIARWAFUDESA 200
QY 188 PLSSGSGDGTGTVDQGLVALPDKSTGLANPGQFAGYTGAAESPTSVLLINHLGHIE 247
Db 201 PLETGSMDSVTGFIADGLLQALGAATTCGLKDAVQFKGSGEAAKPAATILLKNGLHTE 260
QY 248 ILDPESQVGTDRAGVKDVLIESATTTIMDFEDSVAADAAVKVLYGRNWGLKNGDLA 307
Db 261 IVIDPSTEIGKSORAGISDVLIESATTTIMDCEDSVAADAAVKVLYGNWGLMGRDLT 320
QY 308 AAVDKDGTAFRLVLRNDRNTAPGGQFTLPGRSLMFRVNVGHMTNDAIVDTDGSSEVPE 367
Db 321 ENVSKGNTTRRLNPDYRYTAPDGSALTLPGRSLMFRVNVGHMTNPAILDGRDVE 380
QY 368 GINDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVL 427
Db 381 GINDAVVTALIALYDVGS---GRRQNSRAGSMYVVKPKMHGPEEVAFAFANEIPARVENLV 437
QY 428 GLPONTWKIGIMDEERTTVNLKACIKAAADRVVFNFTGFLDRTGDEIHTSMEAGPMVK 487
Db 438 GMAPNTWKIGIMDEERTTVNLKESIRAKDRVFNFTGFLDRTGDEIHTSMEAGPMVK 497
QY 488 GTWKSQPWILLAYEDHNVDAGLAAGFGSRAQVKGKMTMTLMADMVETKIAQPRAGASTA 547
Db 498 GDMKQAIAAYENWNVDIGLECGLSGHAQIGKGMAMPDLMAMLEQKIAHPKAGANTA 557
QY 548 WPSPTAATHALHYHOVDVAAVQOGLAGKRATIEQLTIPLAKELAWAPDEIREVDN 607
Db 558 WPSPTAATHATHYHKVDVAAVQOGLKSRGAKLSILSVVAPRPNTPBEIORELDN 617
QY 608 NCOSILGYVVRWVDQGVGCSKVPDIHDVLMEDRATLRSSQLLANLWLRHGVITTSADVRA 667
Db 618 NAQILGYVVRWVDQGVGCSKVPDINNGLMEDRATLRISAOHMANLWLRHGVVTEAIIK 677
QY 668 SLERMAPLVRQAGDVAYRPMAPNFDSDIAFLAAQELILSGAQONGYTEPILHRRRRE 727
Db 678 TMKRMVAWDTQAGDPAYLPNASDFGSAFAAVALVLRKREQNGYTEPVLHRRRLE 737
QY 728 FKAAA 733
Db 738 LKAKQA 743

RESULT 6

AF3299
malate synthase (EC 4.1.3.2) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AF3299
R;Delvechio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goldsman, E.; Selkov, B.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3299
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-728 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51561.1; PID:gi7982281; GSPDB:GN00190
A:Experimental source: strain 16M
A:Gene: BMEI0380
A:Map position: I
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase
Query Match 38.2%; Score 2318; DB 2; Length 728;
Best Local Similarity 61.8%; Pred. No. 2.7e-131;
Matches 444; Conservative 97; Mismatches 174; Indels 4; Gaps 2;

RESULT 7

140715
malate synthase (EC 4.1.3.2) - Corynebacterium glutamicum
C:Species: Corynebacterium glutamicum
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 05-May-2000
C:Accession: I40715; I40836
R;Reinscheid, D.J.; Eikmanns, B.J.; Sahm, H.
Microbiology 140, 3099-3108, 1994
A:Title: Malate synthase from Corynebacterium glutamicum: sequence analysis of the gene
A:Reference number: I40715; MUID:95111631; PMID:7812449
A:Accession: I40715
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-739 <RES>
A:Cross-references: EMBL:X78491; NID:g530011; PIDN:CAA55243.1; PID:g530012
R;Lee, H.
J. Microbiol. Biotechnol. 4, 256-263, 1994
A:Title: Molecular characterization of aceB, a gene encoding malate synthase in Corynebacterium
A:Reference number: I40836
A:Accession: I40836
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-739 <RE2>
A:Cross-references: GB:I27123; NID:g853573; PIDN:AAA68074.1; PID:g853574

C;Genetics:
A;Gene: aceB
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 37.8%; Score 2289; DB 2; Length 739;
Best Local Similarity 61.3%; Pred. No. 1 5e-129;
Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;

Qy	9	TRVSVGNRIARVLVDFVNNEALPGTDDIDPSFWAGVQKVVADLTLPQNALINARDELQ	68
Db	20	TERVDAGGMQAVKLVDFVTEAVLPRVGVDAERFWSGFAAIARDLTLPNRRELIARDELQ	79
Qy	69	AQIDKWHRRVIEPIEDMDAYRQLTEIGYLLPEDDFTITTSVGDAEITTTAGPQLAVVP	128
Db	80	MLIDDYHRNY-STIQEAYEDFLKEIGLVPEPEAEINTQNVDTETISTAGPQLAVVF	138
Qy	129	LNARFALNANARWGSGLYDALYCTDVPETDGAEGKPTYNKVRGDKVIAYARKFLDSDVP	188
Db	139	LNARFALNANARWGSGLYDALYCTNAIPETDGAEGKPEYNVRGQKVIEMGREFLDSVVP	198
Qy	189	LSSGSGFDATGFFVQGGOLVVALPDKSTGLANFGQAGYTGAAESTSVLLINHGHLIEL	248
Db	199	LDGASHADVEKYNITDGLAAHIGDSVYRKRESYRGFTGNFLDPEAILLETNGHLIEL	258
Qy	249	LIDPESQVGTDRAGVKDVILESAITIMDFEDSVAADAAADKVLGYRNWLGINKGDLAA	308
Db	259	QIDPVPHPGKADTKGLKDIVLESAITIMDFEDSVAADAEADTKLOYSNWFLGNTGELKE	318
Qy	309	AVDKDGTAFVLRLNRDRNYTAPGGGFTLPGRSLMFRNVNGHLMTDAIVDDTDCGSVFEG	368
Db	319	EMSKNGRIETRELNRDRVYIGRNGTELVLHGRSLLFVRNVGHLMQNPSIL-IDGEEIFEG	377
Qy	369	IMDALFTGLTAIHGLKASDVNGPLINSRTGSIYIVKPKHGPAAVFTCELFSRVEDVLG	428
Db	378	IMDAVLTTVCALPGIAPQN---KWRNSRKSGSIYIVKPKHGPAAVFTNELFSRVEDLLD	434
Qy	429	LPONTWKIGIMDBERRITVNLKACIKAAADRVVFINTEGLDRTGDSBIHISMEAGPMVRKG	488
Db	435	LPHTLLKVGVMDBERRITSVNLDASIMEVADRLAFINTGFLDRTGDSBIHISMEAGAMVRKA	494
Qy	489	TKMSQPMILAYEDPHNVDDAGLAAGFSGRAOVGKGMWMTMLMADWVETKTAQPRAGASTAW	548
Db	495	DMGTAPWKQAYENNVDDAGIQRLGPKAQIGKGMWMTMLMADWVETKTAQPRAGANTAW	554
Qy	549	VPSPPTAATLHALHYQVDVAAVQOGL-AGKRRATIEQLTITPLAKELAWAPDIREVDN	607
Db	555	VPSPFTGATLHATHYHLVDVFKVQDELRAAGRRDSLNRILTIPTAPNTNSEEKKEEMDN	614
Qy	608	NCOSILGYVVRVDDQVGCCKVPDIHDVALMEDPATLRISQQLANWLHGVITSDVCA	667
Db	615	NCOSILGYVVRVVEHGVGCCKVPDIHDIDMEDPATLRISQQLANWLHGVITSDVCA	674
Qy	668	SLERMAPLVDRQNAVGVAYRPMAPNPTDDSIATFLAAQELILISGAQQPQNGVTEPILHRRRE	727
Db	675	SLERMAVVVDKQAGDEAYRDMAPNYDASLAFQAANDLIFEKTKSPGYTEPILHARRRE	734
Qy	728	FKAR 731	
Db	735	FKAR 738	

RESULT 8
E83316
malate synthase BH2133 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: E83916
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: E83916
A;Status: preliminary

[illegible]

A;Reference number: S51788; MUID:95010032; PMID:7925370

A;Accession: S51788

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-723 <MOL>

A;Cross-references: EMBL:X74547; NID:G517246; PIDN:CAAS2639.1; PID:G517247

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of *Escherichia coli* K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: F65083

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-723 <BLAT>

A;Cross-references: GB:AE000380; GB:U00096; NID:G1789344; PIDN:AACT6012.1; PID:G1789348;

A;Experimental source: strain K-12, substrain MGI655

C;Genetics:

A;Gene: *glcB*

C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 35.5%; Score 2154.5; DB 2; Length 723;

Best Local Similarity 58.0%; Pred. No. 1.8e-121;

Matches 419; Conservative 115; Mismatches 174; Indels 15; Gaps 5;

Qy 12 VSVGNRIARVLYDFVNNALPGTDIDPDPSFWAGVDKVVADLTPOQALLNARDLQAI 71

Db 5 ITQSLRIDANFRFVDEEVLPGTGLDAAAFWENFDEIVHDLAPENRQLLARDRIQAAL 64

Qy 72 DKHRRRVIPIDMDAYRQPLTIGVILLPEPDFTTTSGVDABITTTAGPOLVVPVUNA 131

Db 65 DEWHRNPGPVKDKAAKYSFLRELGLVLPQPERVTEITGIDSEITSAGPOLVWPANNA 124

Qy 132 RFALNANARWGLYDALYGVDPVPTDGAKEGPTNKYRGDKVAVAKFDDSDVPLSS 191

Db 125 RYALNANARWGLYDALYGVDSIIPOGANVSG--YDPOGGEQVIAWVRFDESILPLEN 182

Qy 192 GSGGDTAGTVQDGLVVALPD-KSTGLANPGFAGYTGAAEPSVTLINHLHIEILI 250

Db 183 GSYQDVAFVAVKQLRIQKNGKETTLPFAQFVGVRGDAAPTCLLKNKGLHIELQI 242

Qy 251 DPSOVTGTDRAQVKDVIRESAITTTIMDFEDSVAADKVLGVNMLGLNKGDLAAAV 310

Db 243 DANGRIKDDPAHINDVIVAAITLTDCEDSVAADKVLGLYKLLNGLNKGDLAAAV 302

Qy 311 DKDGTAPLRVINDRYNTPAGGGQFTLPGRSLMFVENVGHMTNDAIVDTDGESEVFEGIM 370

Db 303 EKNGRQIVRKINDRRHYTAADGSEISLHGKSLFIRNVGHMTIPVWDSENEIPEGIL 362

Qy 371 DALFTGLIATHGKASDVNGPLINSGTGIYVKKPMHGPABVAFTCELPFRVEDVLGP 430

Db 363 DGWMTGAIALYDLKVK-----NSRTGSYVIVKPKMHGPQEVAFANKLFTRIETMLGNA 416

Qy 431 QNTWKIGIMDEERTTNLKAACAKAAADRVVFTINTGFLDRTGDEIHTSMEAGPMVRKGTM 490

Db 417 PNTLKGIMDEERTSILNRSCTAQRNRFVINTGFLDRTGDMHVSWEAGPMLRKQM 476

Qy 491 KSPFWILAYEDHNVDAAGLAFSGRAQVKGKMWMTVELMADVETKIAOPRAGASTAVVP 550

Db 477 KSTFWIKAVERNVLSGLFCGLRGKQAIKGKMWAMPDLNADMYSQKGDQLRAGANTAVVP 536

Qy 551 SPTAATLHALHYHOVDVAAVQOGLA-----GKRATIEOLLTPIAKELAWAPDIREEV 605

Db 537 SPTAATLHALHYHOTVNVQSVQANIAQTEFWAFEPILLDDLLTIPVAENANNSAQIQEL 596

Qy 606 DNNQCISILGVYRWVQGVGCKSVPIHDVAMMEDRATLRISQLLANWLHGVITSADV 665

Db 597 DNNVQGITLVVWRVWQGVGCKSVPIHDVAMMEDRATLRISQHLANWLHGVITKEQV 656

Qy 666 RASLERMAPLVDRONAGDVAVRPMANFDDSTAFLLAAQELILSGAQOPNGYETPILHR-R 724

Db 657 QASLENNAKVQDQNGDPAYRPMAGNFANSCAFKAASDLIFLGKVPQNGYETPILHAWR 716

Qy 725 RRE 727

Db 717 LRE 719

RESULT 10

F70794

probable ask protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C;Accession: F70794

R;Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.B.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70794

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-421 <COL>

A;Cross-references: GB:AL022121; GB:AL123456; NID:G3261559; PIDN:CAA18031.1; PID:G2960133

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: ask

C;Superfamily: aspartate kinase; aspartate kinase homology

F;3-406/Domain: aspartate kinase homology <DKI>

Query Match 26.7%; Score 1615.5; DB 2; Length 421;

Best Local Similarity 85.1%; Pred. No. 1.8e-89;

Matches 338; Conservative 7; Mismatches 35; Indels 17; Gaps 3;

Qy 749 VQYGGSSVADAEIRRVAEIRVATKQGNVWVVSAMGDTTDDLDDLAQQVCPAPPPR 808

Db 5 VQYGGSSVADAEIRRVAEIRVATKQGNVWVVSAMGDTTDDLDDLAQQVCPAPPPR 64

Qy 809 ELDMLLAGRISNALVMAIESLGAHARSFTQAGVITTTGHNAKIIDVTPGRLQTA 868

Db 65 ELDMLLAGRISNALVMAIESLGAHARSFTQAGVITTTGHNAKIIDVTPGRLQTA 124

Qy 869 LEEGRVVLVAGFQVSDTKDVTTLGEGGSDTTAVAMAALGADVCEIYTDVGIISADP 928

Db 125 LEEGRVVLVAGFQVSDTKDVTTLGEGGSDTTAVAMAALGADVCEIYTDVGIISADP 184

Qy 929 RIVENARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRGTVVVGS 988

Db 185 RIVENARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRGTVVVGS 244

Qy 989 IKQVMEDEPILTVGAHRSSEAKVTIUGLPDIPGYAAKVFRAVRRRQHRHGAEE----- 1043

Db 245 IKQVMEDEPILTVGAHRSSEAKVTIUGLPDIPGYAAKVFRAV-----DADVNIDM 295

Qy 1044 RLQGRGQDRHHLHLLPQTS---GPPPKWNTWRTSRASQTLLYDDHIGKVLIGAGMR 1100

Db 296 VLQNVSKVEDGKTDITFTCSRDPVGAPEKLSLRNEIGFSQLLYDDHIGKVLIGAGMR 355

Qy 1101 SHPQVTTATFEALAAVGVNIELISTSDQSRCCAAT 1137

Db 356 SHPQVTTATFEALAAVGVNIELISTSEIRISVLCRDT 392

RESULT 11

GB7199

aspartokinase [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 14-Dec-2001

C;Accession: GB7199

R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hol

R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq

A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: G87199
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL450380; NID:G13093941; PIDN:CAC31839.1; GSPDB:GN00147
C:Genetics:
A:Gene: ask
C:Superfamily: aspartate kinase; aspartate kinase homology

Query Match 25.5%; Score 1545.5; DB 2; Length 421;
Best Local Similarity 80.4%; Pred. No. 2.9e-85;
Matches 319; Conservative 18; Mismatches 43; Indels 17; Gaps 3;

QY 749 VQKYGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQVCPAPP 808
Db 5 VQKYGSSVADAEIRIRVAERIVQTKQNDVVVVVSAMGDTTDLDDLAQVCPAPP 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTGRLQTA 868
Db 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTGRLQTA 124

QY 869 LEEGRVVLVAGFGQVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGFSADP 928
Db 125 LDEGRVVLVAGFGQVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGFSADP 184

QY 929 RIVNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 988
Db 185 RIVNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 244

QY 989 IKDVPMDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE----- 1043
Db 245 IKDVPMDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE-----DADVNIDM 295

QY 1044 RLQGRGRQRHLLH-----LPQSGPPPKWNRSETRSASTQLLYDDHIGKVSLLGAG 1100
Db 296 VLQNVSKVEDGKTDITFTCSRDSPGPIAVAKLSLDEIGFTQLLYDDHIGKVSLLGAG 355

QY 1101 SPFGVTATFCEALAAVGVNIELISTSEDRSCCAAT 1137
Db 356 SPFGVTATFCEALAAVGVNIELISTSEIRISVLCRDT 392

RESULT 12
S42422
aspartate kinase (EC 2.7.2.4) alpha chain - Mycobacterium smegmatis
N:Alternate names: aspartokinase alpha chain
C:Species: Mycobacterium smegmatis
C:Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 18-Jun-1999
C:Accession: S42422; S31801
R:Cirillo, J.D.; Weisbrod, T.R.; Pascopella, L.; Bloom, B.R.; Jacobs Jr., W.R.
Mol. Microbiol. 11, 629-639, 1994
A:Title: Isolation and characterization of the aspartokinase and aspartate semialdehyde
A:Reference number: S42421; MUID:94254720; PMID:7910936
A:Accession: S42422
A:Molecule type: DNA
A:Residues: 1-421 <CIR>
A:Cross-references: EMBL:Z17372; NID:G44506; PIDN:CAA78984.1; PID:G581352
A>Note: the authors translated the initiation codon GTG for residue 1 as val
C:Genetics:
A:Start codon: GTG
C:Superfamily: aspartate kinase; aspartate kinase homology
C:Keywords: alternative initiators; phosphotransferase
F:1-421/Product: aspartate kinase alpha chain #status predicted <ASA>
F:3-406/Domain: aspartate kinase homology <DKI>
F:250-421/Product: aspartate kinase beta chain #status predicted <ASB>

Query Match 24.4%; Score 1478.5; DB 2; Length 421;
Best Local Similarity 77.6%; Pred. No. 3.1e-81;
Matches 302; Conservative 31; Mismatches 35; Indels 21; Gaps 4;

QY 749 VQKYGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQVCPAPP 808

Db 5 VQKYGSSVADAEIRIRVAERIVETKAGNDVVVVVSAMGDTTDDLLDLARQVSPAPP 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTGRLQTA 868
Db 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTGRLQTA 124

QY 869 LEEGRVVLVAGFGQVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGFSADP 928
Db 125 LDEGRVVLVAGFGQVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGFSADP 184

QY 929 RIVNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 988
Db 185 RIVNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 244

QY 989 IKDVPMDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE----- 1043
Db 245 IEDIPMEDALITGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE-----ADVNDM 295

QY 1044 RLQ-----RGRDRHLLHLLPQSGPPPKWNRSETRSASTQLLYDDHIGKVSLLGAG 1098
Db 296 VLQNVSKIEDGKTD--ITFTCARDNGPRAVEKLKALKSEIGFQVLYDDHIGKVSLLGAG 353

QY 1099 MRSPGVTATFCEALAAVGVNIELISTSE 1127
Db 354 MRSPGVTATFCEALAEAGINIDLISTSE 382

RESULT 13
S15276
aspartate kinase (EC 2.7.2.4) alpha chain - Corynebacterium glutamicum
N:Alternate names: aspartokinase alpha chain
C:Species: Corynebacterium glutamicum
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 18-Jun-1999
C:Accession: S15276; S49977
R:Kalinowski, J.; Cremer, J.; Bachmann, B.; Eggeling, L.; Sahm, H.; Puhler, A.
Mol. Microbiol. 5, 1197-1204, 1991
A:Title: Genetic and biochemical analysis of the aspartokinase from Corynebacterium glut
A:Reference number: S15276; MUID:92065816; PMID:1956296
A:Accession: S15276
A:Molecule type: DNA
A:Residues: 1-421 <PUE>
A:Cross-references: EMBL:X57226; NID:G40509; PIDN:CAA40502.1; PID:G580983
R:Experimental source: ATCC 13032
R:Serebrjanski, I.; Wojcik, F.; Reyes, O.; Leblon, G.
submitted to the EMBL Data Library, November 1994
A:Description: Two loci of Corynebacterium glutamicum ATCC17965 that complement Escheric
A:Reference number: S49977
A:Accession: S49977
A:Molecule type: DNA
A:Residues: 329-421 <SER>
A:Cross-references: EMBL:X82928; NID:G599716; PIDN:CAA58100.1; PID:G599717
R:Kalinowski, J.; Bachmann, B.; Thierbach, G.; Puhler, A.
Mol. Gen. Genet. 224, 317-324, 1990
A:Title: Aspartokinase genes lysC-alpha and lysC-beta overlap and are adjacent to the as
A:Reference number: S12250; MUID:91094767; PMID:1980002
A:Contents: annotation
C:Genetics:
A:Gene: lysC-alpha
A:Start codon: GTG
C:Superfamily: aspartate kinase; aspartate kinase homology
C:Keywords: alternative initiators; phosphotransferase
F:3-406/Domain: aspartate kinase homology <DKI>

Query Match 20.3%; Score 1229.5; DB 2; Length 421;
Best Local Similarity 64.3%; Pred. No. 2.7e-55;
Matches 252; Conservative 55; Mismatches 58; Indels 27; Gaps 5;

QY 749 VQKYGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPP 808
Db 5 VQKYGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLLDLAQVCPAPP 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTGRLQTA 868

Db 65 EMDMLTAGERISNALVMAIESLGAFAQFTSQAGVLTTERHGNARIVDVTGVRREA 124
Qy 869 LEGRVVLVAGFGVGSQDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 928
Db 125 LDEKICIVAGFGVGNKTRDTVTTLGRGSGDTTAVAMAAALGADVCEIYSDVDGVYADP 184
Qy 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVGS 988
Db 185 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVGS 244
Qy 989 IKDVPMDPILTVGAHDSRSEAKVTIVGLPIPGYAAKVFRAVA- - - - -RRRQHR 1038
Db 245 MEDIFVEEAVLTGATDKSEAKVTIVGLISDKPGEAAKVFRALADAENIDMVLQNVSSYE 304
Qy 1039 HGAARLQGRQD-RHHLHP--QTSGPPPPWNTTRSETRSASTQLLYDDHIGKVSULI 1095
Db 305 DGTDTITFTCFRSDGRAMEILKLQVQ- - - - -TNVLYDDQGVKSLV 350
Qy 1096 GAGVRSHPGVYATFCEALAAVGVNIELISTSE 1127
Db 351 GAGVRSHPGVYATFCEALAAVGVNIELISTSE 382

RESULT 14
T35383
probable aspartokinase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 31-Jan-2000
C:Accession: T35383
R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21576
A:Accession: T35383
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <MUR>
A:Cross-references: EMBL:AL079348; PIDN:CAB45482.1; GSPDB:GN00070; SC066T3.26c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: ask; SC066T3.26c
C:Superfamily: aspartate kinase; aspartate kinase homology

Query Match 18.8%; Score 1141; DB 2; Length 425;
Best Local Similarity 61.6%; Pred. No. 5.7e-61;
Matches 239; Conservative 54; Mismatches 79; Indels 16; Gaps 4;

Qy 749 VQYGGSSVADAEIRRVARRIVATKQGNVAVVVSAMGDTTDDLLDLAQVCPAPP 808
Db 5 VQYGGSSVADAEIRRVARRIVATKQGNVAVVVSAMGDTTDDLLDLAQVCPAPP 64
Qy 809 ELDMLTAGERISNALVMAIESLGAFAQFTSQAGVLTTERHGNARIVDVTGVRREA 868
Db 65 ELDVAVSTGEQVTTALISMLIKKGVPAVSYTGNQVRLTDSATKARILHIDTHIRAD 124
Qy 869 LEGRVVLVAGFGVGSQDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 928
Db 125 LKAGRVVAVAGFGV-DGNGNITTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 183
Qy 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVGS 988
Db 184 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVGS 243
Qy 989 IKDVPMDPILTVGAHDSRSEAKVTIVGLPIPGYAAKVFRAVA- - - - -1036
Db 244 DEESSMEQPIISGIAFNDRDEAKLTIRGVPTPGVAFKILGPISAANVVDVMIQNVAHN 303
Qy 1037 - - - - -HRHGAERLQGRQDRHHLHPQTSQAGVLTTERHGNARIVDVTGVRREA 1088
Db 304 TDTFTVHRYNDYLNAL- - - - -ILQTA- - - - -ANIGAREAGDTN 340
Qy 1089 IGVSLICAGVRSHPGVYATFCEALAAVGVNIELISTSE 1127
Db 341 IAKVSIQVGVRSAGVRSHPGVYATFCEALAAVGVNIELISTSE 379

Search completed: November 21, 2003, 16:09:56
Job time : 26.4499 secs

Db 358 KSNFGVTDFTALSDAGVNIELISTSE 385

RESULT 15
C83531
aspartate kinase alpha and beta chain PA0904 [imported] - Pseudomonas aeruginosa (strain
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 14-Dec-2001
C:Accession: C83531
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br-
edman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathob-
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: C83531
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <STO>
A:Cross-references: GB:AE004525; GB:AE004091; NID:g9946805; PIDN:AAG04293.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: lysC; PA0904
C:Superfamily: aspartate kinase; aspartate kinase homology

Query Match 13.6%; Score 823; DB 2; Length 412;
Best Local Similarity 42.6%; Pred. No. 6.8e-42;
Matches 170; Conservative 80; Mismatches 105; Indels 44; Gaps 4;

Qy 749 VQYGGSSVADAEIRRVARRIVATKQGNVAVVVSAMGDTTDDLLDLAQVCPAPP 808
Db 5 VQYGGSSVADAEIRRVARRIVATKQGNVAVVVSAMGDTTDDLLDLAQVCPAPP 64
Qy 809 ELDMLTAGERISNALVMAIESLGAFAQFTSQAGVLTTERHGNARIVDVTGVRREA 868
Db 65 ELDVAVSTGEQVTTALISMLIKKGVPAVSYTGNQVRLTDSATKARILHIDTHIRAD 124
Qy 869 LEGRVVLVAGFGVGSQDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 928
Db 125 LKAGRVVAVAGFGV-DGNGNITTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 183
Qy 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVGS 988
Db 184 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVGS 243
Qy 989 IKDVPMDPILTVGAHDSRSEAKVTIVGLPIPGYAAKVFRAVA- - - - -1036
Db 244 DEESSMEQPIISGIAFNDRDEAKLTIRGVPTPGVAFKILGPISAANVVDVMIQNVAHN 303
Qy 1037 - - - - -HRHGAERLQGRQDRHHLHPQTSQAGVLTTERHGNARIVDVTGVRREA 1088
Db 304 TDTFTVHRYNDYLNAL- - - - -ILQTA- - - - -ANIGAREAGDTN 340
Qy 1089 IGVSLICAGVRSHPGVYATFCEALAAVGVNIELISTSE 1127
Db 341 IAKVSIQVGVRSAGVRSHPGVYATFCEALAAVGVNIELISTSE 379

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 13.8195 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-52

Perfect score: 5061

Sequence: 1 MQHHHHHTDRSVGNURIA.....SAATRRPCTRGDRGWACQ 1172

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	62.9	741	1 MASZ_MYCTU	Q50596 Mycobacteri
2	3117.5	51.4	731	1 MASZ_MYCLE	Q32913 Mycobacteri
3	2634	43.5	724	1 MASZ_RHOFA	Q9a55 rhodococcus
4	2569	42.4	725	1 MASZ_PSEAE	Q91636 pseudomonas
5	2541	41.9	725	1 MASZ_PSEPK	Q88qx8 pseudomonas
6	2497	41.2	725	1 MASZ_PSEEL	Q05137 pseudomonas
7	2489	41.1	725	1 MAZI_PSESM	Q88ab2 pseudomonas
8	2416	39.9	721	1 MASZ_RHILO	Q98dk4 rhizobium l
9	2380	39.3	731	1 MASZ_AGRF5	Q8uj85 agrobacteri
10	2367.5	39.1	723	1 MASZ_RHILV	Q937w7 rhizobium l
11	2351.5	38.8	723	1 MASZ_RHIME	Q92ta4 rhizobium m
12	2318	38.2	728	1 MASZ_BRUME	Q8yir3 brucella me
13	2313	38.2	728	1 MASZ_BRUSU	Q8fz50 brucella su
14	2289	37.8	738	1 MASZ_CORGL	P42450 corynebacte
15	2273	37.5	741	1 MASZ_COREF	Q8fmb3 corynebacte
16	2272.5	37.5	721	1 MASZ_BRAVA	Q89ue3 bradyrhizob
17	2258	37.3	727	1 MASZ_BACHD	Q9kb03 bacillus ha
18	2215	36.5	725	1 MAZ2_PSESM	Q87272 pseudomonas
19	2168.5	35.8	722	1 MASZ_EC0L6	Q8fdn6 escherichia
20	2156.5	35.6	722	1 MASZ_SHPLI	P96633 shigella fl
21	2154.5	35.5	722	1 MASZ_EC0LI	P37300 escherichia
22	1615.5	26.7	421	1 AK MYCTU	P37048 mycobacteri
23	1478.5	24.4	421	1 AK MYCSM	P41403 mycobacteri
24	1248.5	20.6	421	1 AK COREF	Q8qnl1 corynebacte
25	1224.5	20.2	421	1 AK CORGL	P36512 corynebacte
26	1216.5	20.1	421	1 AK CORFL	P41398 corynebacte
27	823	13.6	412	1 AK_PSEAE	Q89077 pseudomonas
28	819	13.5	405	1 AK_TRETH	P97151 thermus the
29	785	13.0	411	1 AK2_BACST	Q59229 bacillus sp
30	778	12.8	407	1 AK2_BACST	P53553 bacillus st
31	734.5	12.1	415	1 AK_AOUAE	Q87221 aquifex aeo
32	721	11.9	405	1 AK_HELPJ	Q8zjz7 helicobacte
33	712.5	11.8	408	1 AK2_BACSU	P68495 bacillus su

RESULT 1

MASZ_MYCTU	711	11.7	405	1	AK_HELPY	025827 helicobacte
34	532	8.8	404	1	AK1_BACSU	Q04795 bacillus su
35	492	8.1	473	1	AK_VETJA	Q57991 methanococ
36	480.5	7.9	446	1	AK_RICPR	Q92a17 rickettsia
37	369.5	6.1	819	1	AKIH_SERMA	P27725 serattia ma
38	364.5	6.0	820	1	AKIH_EC0LI	P00561 escherichia
39	324	5.3	921	1	AKH_DAUCA	P37142 daucus caro
40	319.5	5.3	815	1	AKH_HABIN	P44505 haemophilus
41	310.5	5.1	449	1	AK3_EC0LI	P08560 escherichia
42	308.5	5.1	454	1	AK3_BACSU	P94417 bacillus su
43	285	4.7	816	1	AKH_BUCAI	P57290 buchnera ap
44	282	4.7	431	1	AK_CHLTR	O84367 chlamydia t
45						

ALIGNMENTS

STANDARD; PRT; 741 AA.

Q50596;

01-NOV-1997 (Rel. 35, Created)

01-NOV-1997 (Rel. 35, Last sequence update)

15-SEP-2003 (Rel. 42, Last annotation update)

Malate synthase G (EC 2.3.3.9).

GN GUCB OR RV1837C OR MT1885 OR MTCY1A11.06.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekoa F.,

Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

"Deciphering the biology of Mycobacterium tuberculosis from the

complete genome sequence."

Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,

Bishai W.;

"Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains."

Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +

CoA.

CC -1- PATHWAY: Glyoxylate bypass; second step.

CC -1- SUBUNIT: Monomer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- SIMILARITY: Belongs to the malate synthase family. GLOS subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to license@isb-sib.ch.

CC -----

DR EMBL; Z78020; CAB01465.1; --

DR EMBL; AE007047; AAK46156.1; --


```

DR PIR; F70722; F70722.
DR PDB; 1N81; 18-DEC-02.
DR PDB; 1N8W; 18-DEC-02.
DR TIGR; MT1885; -.
DR Tuberculist; Rv1837c; -.
DR HAMAP; MF_00641; -.
DR InterPro; IPR001465; Malate_synthase.
DR InterPro; IPR006253; Malate_synthase.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate_synthase; 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome; 3D-structure.
FT ACT_SITE 339 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 633 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 741 AA; 80403 MW; A92F54E0F8B7C64 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e-205;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
DB 2 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 61
QY 69 AQIDKWHRRVIEPIIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 128
DB 62 AQIDKWHRRVIEPIIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 121
QY 129 LNARFALNANRWGSLYDALYGTVDIPETDGAEGKPTNKVRGKVAYARFLLDDSV 188
DB 122 LNARFALNANRWGSLYDALYGTVDIPETDGAEGKPTNKVRGKVAYARFLLDDSV 181
QY 189 LSSGSGFGDATGFTVQDQQLVWALPDKSTGLANGPQAGTGAESPTSVLLINHLGHEI 248
DB 182 LSSGSGFGDATGFTVQDQQLVWALPDKSTGLANGPQAGTGAESPTSVLLINHLGHEI 241
QY 249 LIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAADVAADKVLGYRNWGLNGKDLAA 308
DB 242 LIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAADVAADKVLGYRNWGLNGKDLAA 301
QY 309 AVDKDGTAFILVLRNDRNTAPGGGFTLPGSLMFRVNVGHLMNDATVDTDSEVFE 368
DB 302 AVDKDGTAFILVLRNDRNTAPGGGFTLPGSLMFRVNVGHLMNDATVDTDSEVFE 361
QY 369 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMHPAEVAFTCELSFVEDVLG 428
DB 362 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMHPAEVAFTCELSFVEDVLG 421
QY 429 LPONTWKIGIMDEERTTNLAKIAAADRVVINTGFLDRTGDEIHTSMAGPMVRKG 488
DB 422 LPONTWKIGIMDEERTTNLAKIAAADRVVINTGFLDRTGDEIHTSMAGPMVRKG 481
QY 489 TMKSQWILLAYEDHNVDAGLAFSGRAQVQKGMWNTMELMADWVETKIAQPRAGASTAW 548
DB 482 TMKSQWILLAYEDHNVDAGLAFSGRAQVQKGMWNTMELMADWVETKIAQPRAGASTAW 541
QY 549 VPSPTAATHALHYHVDVAAVOOGLAGKRRTATIEQLTIPIAKELANAPDSIREVDNN 608
DB 542 VPSPTAATHALHYHVDVAAVOOGLAGKRRTATIEQLTIPIAKELANAPDSIREVDNN 601
QY 609 CQSILGVVVRWVDQVGCCKVPDIHVALMDEBRATLRISSQLLANWLRHGVITSDVRAS 668
DB 602 CQSILGVVVRWVDQVGCCKVPDIHVALMDEBRATLRISSQLLANWLRHGVITSDVRAS 661
QY 669 LERMAPLVDRQAGVAYRPMAPNFDDSTAFIAAQELILSGAQQNGVTEPILHRRREF 728
DB 662 LERMAPLVDRQAGVAYRPMAPNFDDSTAFIAAQELILSGAQQNGVTEPILHRRREF 721
QY 729 KARAEEKPAPSDRAGDAAAR 748
DB 722 KARAEEKPAPSDRAGDAAAR 741

```

```

RESULT 2
MASZ MYCLE
ID MASZ MYCLE STANDARD; PRT; 731 AA.
AC O32913;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR ML2069 OR MLCBL788.27.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RA MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hanlin N.,
RA Holroyd S., Hornsby I., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 408:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GICB subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL008609; CAAL5459.1; -.
CC EMBL; AL583924; CAC31024.1; -.
CC PIR; T44752; T44752.
CC HSP; P37330; 1D8C.
CC Leprosa; ML2069; -.
CC HAMAP; MF_00641; -.
CC InterPro; IPR001465; Malate_synthase.
CC InterPro; IPR006253; Malate_synthase.
CC Pfam; PF01274; Malate_synthase; 1.
CC TIGRFAMs; TIGR01345; malate_synthase; 1.
CC Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 638 638 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 731 AA; 80141 MW; 3878CADA45DB416C CRC64;

Query Match
Best Local Similarity 51.4%; Score 3117.5; DB 1; Length 731;
Matches 597; Conservative 54; Mismatches 72; Indels 5; Gaps 2;

QY 9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
DB 2 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 61
QY 69 AQIDKWHRRVIEPIIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 128
DB 62 AQIDKWHRRVIEPIIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 121
QY 129 LNARFALNANRWGSLYDALYGTVDIPETDGAEGKPTNKVRGKVAYARFLLDDSV 188
DB 122 LNARFALNANRWGSLYDALYGTVDIPETDGAEGKPTNKVRGKVAYARFLLDDSV 188

```


Db 122 LNARFALNAARWGLYDALYGTDTIPETEGAEGSEYKNKIRGDKVIAYARKEMDQAVP 181
 Qy 139 LSSGSGDATGFTVQDQGVVALPDKSTGLANPCQAGYGAASPT-SVLLINHLGHE 247
 Db 182 LASDWTNATGVSIFDQGLQIAITNSTGLASPEKFGYRNQLRSSWVLLANHGHE 241
 Qy 248 ILIDPESOGTTRAGVKDVILESAITIMDFEDSVAADAAADKVLGRNWLGNKGDLA 307
 Db 242 VLIDPESPIGKTPVGIKDVILESAITIMDFEDSVTAVDADKVRGNWLGNKGDLT 301
 Qy 308 AAVKDGTAFLVRLNDRNYTAPGGGFTLPGRSLMFVRNVGHMTNDIV---DPDGS 363
 Db 302 BEVNKDGKTRVLNADRSTYTPDGGELTLPGRSLLFVRNVGHMTNDIV---DPDGS 361
 Qy 364 EVFEGIMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMHPAEVAFTCELSRV 423
 Db 362 EVFEGIDAVFTGLAAHGLKGTGANGPLINSRTGSIYIVKPKMHPAEVAFTCELSRV 421
 Qy 424 EDVLGLPONTMKGIMDEERTTNLKAICAKAAADRVVFTINTGFLDRTGDEIHTSMEAGP 483
 Db 422 EDVLGLPONTMKGIMDEERTTNLKAICAKAAADRVVFTINTGFLDRTGDEIHTSMEAGP 481
 Qy 484 MVRKGTWKSQWILAYEDHNVDAGLAAGFSGAQVGKGMWMTMELMADMEVETKIAOPRAG 543
 Db 482 MIRKGMKNSTWIKAYEDANVDIGLAAGFGKGAQIGKGMWMTMELMADMEVETKIAOPRAG 541
 Qy 544 ASTAVPSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLTIPLAKELAMAPDEIRE 603
 Db 542 ATTAVPSPPTAATLHAMHYQVDVAAVQOGLTGORRATVDQLTIPLAKLAWAPEIRE 601
 Qy 604 EVDNQCQILGVVRVWDQGVCSKVPDIHDVALMEDRATRISSQLLANWLHGVITSA 663
 Db 602 EVDNQCQILGVVRVWDQGVCSKVPDIHVALMEDRATRISSQLLANWLHGVITSE 661
 Qy 664 DVRASLERMAPLVDRONAGDVAYRPMAPNFDSDTAFLAAQELISGACQNGVTEPIIHR 723
 Db 662 DVRASLERMAPLVDRONAGDVAYRPMAPNFDSDTAFLAAQELISGACQNGVTEPIIHR 721
 Qy 724 RRRBFKAR 731
 Db 722 RRRBFKAQ 729

RESULT 3
 MASZ_RHOFA STANDARD; PRT; 724 AA.
 AC Q9AE55;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Malate synthase G (EC 2.3.3.9).
 GN GLCB OR VICA.
 OS Rhodococcus fascians.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Nocardiaceae; Rhodococcus.
 OX NCBI_TaxID=1828;
 RN [1]_TaxID=1828;
 RP SEQUENCE FROM N.A.
 RC STRAIN=D188;
 RA Verecke D.M., Cornelis K., Van Montagu M., El Jaziri M., Holsters M.,
 RA Goethals K.;
 RT "Characterization of a chromosomal locus that affects pathogenicity in
 Rhodococcus fascians";
 RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
 CoA.
 CC -!- PATHWAY: Glyoxylate bypass; second step.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the malate synthase family. GlsB subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ301559; CAC35701.1; -
 DR HSSP; P37330; 1D8C
 DR HAMAP; MF_00641; 1
 DR InterPro; IPR001465; Malate_synthase.
 DR InterPro; IPR006253; Malate_synthase.
 DR Pfam; PF01274; Malate_synthase; 1
 DR TIGRFAMs; TIGR01345; malate syn G; 1
 KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
 FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
 SQ SEQUENCE 724 AA; 78609 MW; F899FE883890995E CRC64;
 Query Match 43.5%; Score 2634; DB 1; Length 724;
 Best Local Similarity 69.0%; Pred. No. 7.5e-140;
 Matches 499; Conservative 82; Mismatches 138; Indels 4; Gaps 2;
 Qy 9 TDRVSVGNRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTQPONQALLNARDEIQ 68
 Db 2 TDRVQAGQVAKVLFDFVEKEALPGTDIDPDSFAWAGAAVADLAPKPKALLAVRDEIQ 61
 Qy 69 AQIDKMRHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVWPV 128
 Db 62 GKVDAMHGEHAGAEYDRAAYKAFLEIGYLLPEPADFQIHTSGVDTEITTTAGPQLVWPV 121
 Qy 129 LNARFALNAARWGLYDALYGTDTIPETEGAEGSEYKNKIRGDKVIAYARKEMDQAVP 188
 Db 122 LNARFALNAARWGLYDALYGTDTIPETEGAEGSEYKNKIRGDKVIAYARKEMDQAVP 181
 Qy 189 LSSGSGDATGFTVQDQGVVALPDKST-GLANPCQAGYGAASPT-SVLLINHLGHE 247
 Db 182 LSSGSHVGTGYVDDAASLTVLADGSTVGLKDPQLGQYGTPTDPTAILFVHNLGHE 241
 Qy 248 ILIDPESOGTTRAGVKDVILESAITIMDFEDSVAADAAADKVLGRNWLGNKGDLA 307
 Db 242 IQIDPESPIGKTDGAGVKDVILESAITIMDFEDSVAADAAADKVLGRNWLGNKGDLT 301
 Qy 308 AAVKDGTAFLVRLNDRNYTAPGGGFTLPGRSLMFVRNVGHMTNDIVDSDGSEVE 367
 Db 302 BEVSKGKGTFRAMNKDRTYTSVDSSELTGHRSLLFVRNVGHMTSDAILDADGNEVE 361
 Qy 368 GIMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMHPAEVAFTCELSRVEDVL 427
 Db 362 GILDALFTSLAGLSLTDPNV---LSNRTGSLYIVKPKMHPDEVAFTAEIIFGRVEQVL 418
 Qy 428 GLPQNTMKIGIMDEERTTNLKAICAKAAADRVVFTINTGFLDRTGDEIHTSMEAGVNRK 487
 Db 419 GLPTNTLKVIMDEERTTNLKAICAAASERVVFTINTGFLDRTGDEIHTSMEAGVNRK 478
 Qy 488 GTWKSQWILAYEDHNVDAGLAAGFSGAQVGKGMWMTMELMADMEVETKIAOPRAGASTA 547
 Db 479 GANKGKNIAAYEDFNVDIGLAGLQCKAQIGKGMWMTMELMADMEVETKIAOPRAGASTA 538
 Qy 548 WVPSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLTIPLAKELAWAPEIREVDN 607
 Db 539 WVPSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLTIPLAKELAWAPEIREVDN 598
 Qy 608 NCOSILGVVRVWDQGVCSKVPDIHDVALMEDRATRISSQLLANWLHGVITSAADVA 667
 Db 599 NSOSILGVVRVWDQGVCSKVPDIHDVALMEDRATRISSQLLANWLHGVITSAADVA 658
 Qy 668 SLERMAPLVDRONAGDVAYRPMAPNFDSDTAFLAAQELISGACQNGVTEPIIHRRRR 727
 Db 659 SLERMAPLVDRONAGDVAYRPMAPNFDSDTAFLAAQELISGACQNGVTEPIIHRRRR 718
 Qy 728 PKA 730
 Db 719 YKA 721

```

RESULT 4
MASZ_PSEAE STANDARD; PRT; 725 AA.
ID MASZ_PSEAE STANDARD; PRT; 725 AA.
AC Q91636;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR PA0482.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10994043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino B., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Felger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -1- PATHWAY: Glyoxylate bypass; second step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the malate synthase family. GLCB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB004485; AAC03871.1; -.
DR PIR; H83586; H83586.
DR HAMAP; MF 00641; -.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRams; TIGR01345; malate_syn_G; 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 725 AA; 78659 MW; 3669670A9E38D391 CRC64;
Query Match 42.4%; Score 2569; DB 1; Length 725;
Best Local Similarity 69.4%; Pred. No. 3.2e-136;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;
QY 9 TDRVSGNLIARVLDFVNNALPGTDIDPDSFWAGVDKVVADLTPONQALNARDELQ 68
DB 2 TERVQVGLQVAKVLFDFVNNALPGTVGSADFTWGAENVINDLAPKNKALLAKREDELQ 61
QY 69 AQIDKRRHREIVFIDNDAYRQLTEIGYLLPEPDDFTITTSQVDAEITTTAGPOLVVPY 128
DB 62 AKIDGWHQAGAGAHDAVAKFLEIEGYLLPEADFOAGTQVNDDEIARMAPOLVWPV 121
QY 129 LNARFALNANRWGSLYDALYGTVDVTPETDGAEGKPTNKVRGDKVIAAYAKFLDDSDVP 188
DB 122 MNARFALNANRWGSLYDALYGTVDVISEGGAEKKGKNGKVRGDKVIAFAFLDEAP 181

```

```

189 LSSGSGDATGFTYDQGLVVALPKS - TGLANPGFAGYTGAESPTSVLLINHLHIE 247
182 LESGSHVDATSYSVKNGALVVALKNGSETGLKNGAQFLAQGDAAKPQAVLLKHNGLHFE 241
248 ILIDPESQVGTTRAGVKVILSAITTIMDFSDSVAADVDAADKVLGYRWLGLNGKDLA 307
242 IQIDFSSPVGQTDAAAGKVDVLMFAALTIMDCSDSVAADVDAADKVIYRWLGLNGKDLA 301
308 AAVDKDGTAFILVLRNRDNTAPGGGQFTLPGSLMFRVNVGHLMNDALVDTDGSEVPE 367
302 EVSKGGSTTRTNWPNRVVTRADGSLILHGLSLILFVRNVGHLMNDALVDXGNEVPE 361
368 GIMDALFTGLIAHGLKASDVNGPLI - NSRTGSIYIVKPMHGAFAVFTCELSERVED 425
362 GIQDGLFTSLIAIH - - - - - DLNGNTRSKNSRTGSVIVKPMHGPBEEAFTNLSERVED 416
426 VLGLPONTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMV 485
417 VLGLPNTLVKGINMDEERRITVNLKACIKAAKORVVFIINTGFLDRTGDEIHTSMEAGAVV 476
486 RKGTKSQPWILAYEDHNDVAGLAAGSGRAQVGKGMWMTTELMADMVETKIAQPRAGAS 545
477 RKGAMKSEKVI GAYENNVNDVGLATGLQGRAQIGKGMWMPEDLMAAMLEOKIGHPLAGAN 536
546 TAWVPSTAAATLHALHYHQVDVAQVQGLAGKRRATIEQLTTIPLAKELAWADEIREEV 605
537 TAWVPSTAAATLHALHYHKVDVFAQAEALAKRTPASVDILTITPLAPNTNWTAEIKNEV 596
606 DNNQCSILGVYRWVDQGVGCKVPDIHDVALMEDRATLRISQOLLANLWRHGVITSADV 665
597 DNNAGGILGVYRWVDQGVGCKVPDINDVGLMEDRATLRISQOLLANLWRHGVISQEQV 656
666 RASLERMAPLVDRONAGDVAYRMAPNFDSDSIAPLAQELILSGAOPNGYTPILHRRR 725
657 VESLKRMAVVVDRONADSPSYRMAPNFDNDVAFQALELVVEGTQPNGYTPVLHRRR 716
726 REFKA 731
717 REFKA 722
RESULT 5
MASZ_PSPBK STANDARD; PRT; 725 AA.
ID MASZ_PSPBK STANDARD; PRT; 725 AA.
AC Q8QX8;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR P20356.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=150488;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolchay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapfle E., Scanlan D., Tran K., Moazzez A.,
RA Utluerback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Sjepeandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eissen J., Timmis K.N., Dueterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -1- PATHWAY: Glyoxylate bypass; second step.
CC -1- SUBUNIT: Monomer (By similarity).

```


Db 242 IQIDASTPYGQTDAGVKVLMEEAALTITIMDCEDSVAAVDADKVVIVYRNWGLMKGDLA 301
QY 308 AAVDKDGTAFRLVNLDRNYTAPGGQFTLPGRSLMFVRNVGHLMNDIAVTDGSEVPE 367
Db 302 EEVAKGKFTTRTMNPDVYTVGQDVTLHGRLSLLFVRNVGHLMNTDAILDKAGNEVPE 361
QY 368 GIMDALFTGLIAHGLKASDVNGPLNSRTGSITYVKKPHGPAEVAFTCELFSEVEDVL 427
Db 362 GILDGGLTSLAAHSLJNG---NSSRKNRSTGSVYIVKPHGPEEAAFTNELFGRIEDVL 418
QY 428 GLPQNTMKIGIMDEERTTNLAKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRK 487
Db 419 NUPRNLTKVIMDEERTTNLAKACIKAAASERVVINTGFLDRTGDEIHTSMEAGPMVRK 478
QY 488 GTKSPWILAYEDHNVDAGLAFSGRAQVKGKMTTELADWVYKIAOPRAGASTA 547
Db 479 AAMKTEKVIAGVENWVVDIGLSTGLGQRAQIGKGMWAMPDLMAALEQKIAHPLAGANTA 538
QY 548 WPSPTAATLHALHYHVDVAAVQOGLAGKRRTIEQLLTIPLAKELAWAPDEIREVEDN 607
Db 539 WPSPTAALHALHYHVDVFAQRAELAKERASVDDILTIPLAKNTOWSEIEINELDN 598
QY 608 NCOSILGYVVRWVQGVGSKVPDIHDVALMEDRATLRISOLLANWLPKHGVTISADVRA 667
Db 599 NAQILGYVVRWVQGVGSKVPDINDVGLMEDRATLRISQHIANWLRHGVVTDQVME 658
QY 668 SLERMAPLVDQRNAGVAVRPMAPNFDSDIAFLAAQELISGAQPNGYTEPILHRRRRE 727
Db 659 SLKMAPVVDQRNAGDALVRPLAPDPSDNIAQMAVELVIEGTQKPNGYTEPVLHRRRE 718
QY 728 FKAR 731
Db 719 FKAK 722

RESULT 7

MAZI_PSESM STANDARD; PRT; 725 AA.
ID MASI_PSESM
AC Q88AB2;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G 1 (EC 2.3.3.9).
GN GICB1 OR GICB-1 OR PSET00480.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RI "Complete sequence of Pseudomonas syringae";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -1- PATHWAY: Glyoxylate bypass; second step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the malate synthase family. GicB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to a license@isb-sib.ch).
CC -----

DR EMBL; AE016857; AAO54024.1; -.
DR TIGR; PSET00480; -.
DR HAMAP; MF_00641; -. 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 340 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 631 CATALYTIC ACID (BY SIMILARITY).
FT ACT_SITE 631 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 725 AA; 79143 MW; P193264E8083660 CRC64;
Query Match 41.1%; Score 2489; DB 1; Length 725;
Best Local Similarity 65.3%; Pred. No. 9.5e-132;
Matches 473; Conservative 97; Mismatches 150; Indels 4; Gaps 2;
QY 9 TDRYSVGNLRIARVLYDVFVNEALPGTDIDPDSFWAGVDKVVADLTPQOALLNARDELQ 68
Db 2 TEYVQVGDQLVARVLPDFVQNEATPGTVDAGAFWAGADOLIHDLAPKPKALLAQADELQ 61
QY 69 AQIDKWHRRVIFPIDMDAYRQFLTEIGYLLPDPDDFTITTSVDAEITTTAGPQLVVPV 128
Db 62 AQIDAWHQRAGQAHDAAYKAFLEIGYLLPEAADFQITTONVDEBIATMAGPQLVVPV 121
QY 129 LNAFALNANANAGSLYDALYGTDPVETDGAEGKGTYNKVGDKVIAVARFELDDSV 158
Db 122 MNARFALNANANAGSLYDALYGTDAISEGGAEKGGYKVGDKVIAVARFELDDSV 181
QY 189 LSSGSFGDATGFTVQCGQLVVALPDKS-TGLANPGQFAGVTGAESPTSLLINHLHIE 247
Db 182 LAAGSHVDSTAYKLIDGRVLISLKGSGNTGLRDAQLVGFQGDASAPFAVLFKHGLHPE 241
QY 248 ILIDPESQVGTDBAGVKDVILESATTTIMDFDSVAANDAAKVLGYRNWGLNKGDLA 307
Db 242 LQIDAAFPVQGTDPAGVKDIWMEAAUUTTINDCEDSIAAVDADDKVVVYRNWGLMKGDV 301
QY 308 AAVDKDGTAFRLVNLDRNYTAPGGQFTLPGRSLMFVRNVGHLMNDIAVTDGSEVPE 367
Db 302 EESVSKGGETTRTMNPDVYTVGQDVTLHGRLSLLFVRNVGHLMNTDAILDKAGNEVPE 361
QY 368 GIMDALFTGLIAHGLKASDVNGPLNSRTGSITYVKKPHGPAEVAFTCELFSEVEDVL 427
Db 362 GILDGGLTSLAAHSLJNG---NSSRKNRSTGSVYIVKPHGPEEAAFTNELFGRIEQVL 418
QY 428 GLPQNTMKIGIMDEERTTNLAKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRK 487
Db 419 GLPNTLVKIMDEERTTNLAKACIKAAASERVVINTGFLDRTGDEIHTSMEAGPMVRK 478
QY 488 GTKSPWILAYEDHNVDAGLAFSGRAQVKGKMTTELADWVYKIAOPRAGASTA 547
Db 479 AAMKTEKVIAGVENWVVDIGLSTGLGQRAQIGKGMWAMPDLMAALEQKIAHPLAGANTA 538
QY 548 WPSPTAATLHALHYHVDVAAVQOGLAGKRRTIEQLLTIPLAKELAWAPDEIREVEDN 607
Db 539 WPSPTAALHALHYHVDVFAQRAELAKERASVDDILTIPLAKNTOWSEIEINELDN 598
QY 608 NCOSILGYVVRWVQGVGSKVPDIHDVALMEDRATLRISOLLANWLRHGVVITSADVRA 667
Db 599 NAQILGYVVRWVQGVGSKVPDINDVGLMEDRATLRISQHIANWLRHGVVTDQVME 658
QY 668 SLERMAPLVDQRNAGVAVRPMAPNFDSDIAFLAAQELISGAQPNGYTEPILHRRRRE 727
Db 659 SLKMAPVVDQRNAGDALVRPLAPDPSDNIAQMAVELVIEGTQKPNGYTEPVLHRRRRE 718
QY 728 FKAR 731
Db 719 FKAK 722
RESULT 8
MAZI_RHILO STANDARD; PRT; 721 AA.
ID MASI_RHILO
AC Q98DK4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MLR4664.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OX Phyllobacteriaceae; Mesorhizobium.
RN NCBI_TaxID=381;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Katanabe T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shampo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti".
RL DNA Res. 7:331-338(2000).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family, GlsB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003004; BAB51267.1; -.
DR HAMAP; MF_00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; Malate_synth_G.1.
KW Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT SITE 338
FT ACT SITE 629
FT ACT SITE 629
SQ SEQUENCE 721 AA; 78058 MW; 65376311A7E1BDFD CRC64;

Query Match 39.9%; Score 2416; DB 1; Length 721;
Best Local Similarity 64.4%; Pred. No. 1.1e-127;
Matches 467; Conservative 86; Mismatches 164; Indels 8; Gaps 3;

QY 9 TDRVSGNLRIRLVLYDFVWNEALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
Db 2 TDRIEIAGRIAGELHDFVAGEALPGTGTGIADATFWSGSAIVHDLAPKRALKKRAMQ 61

QY 69 AQIDKWHRRVIEPIDMAYRQFLTEIGYLPEDDDFTITTSQGVDAEITTAGPOLVVPV 128
Db 62 ERLDQWYRDNG-APYDMEVYSFLKEIGYLPVEGPAFVSVDNDVPEIYVAGVQLVVPV 120

QY 129 LNARFALNAARWGSGLDLYLGTVDVIEPTDGAEGKPTYNKVRGDKVIAYARKFLDSDVP 188
Db 121 MWARYALNAARWGSGLDLYLGTVDVIEPTDGAEGKGFNPARGAKVIAMAKOFLDOSVP 180

QY 189 LSSGFGDGTFTVDQGLWVALPKSTGLANPQGFAGYTGAASTSVLLINLHGLHIEI 248
Db 181 LTSKGWAGVGLSVAGHAKLKGAGAGGTTADPQGFAGYRGDAANPDPAVLLVKNLHIEI 240

QY 249 LIDPSOYGTTRAGVKVILLESALTITMDPDSVAADADKVLGYRNWGLNKLGLAA 308
Db 241 VDRNNQIGRTDPAGIADVILLESALTITQDCDSVAADQKVYVYRNWGLNKLGLAAE 300

QY 309 AVDDKGTAFRLVNRDRNYTPGGQOFTLPGRSLMFVENVGHMTNDAIVTDGSEVFEG 368
Db 301 BISKAGRSFVKLNADRAYTAPAGQITVPGESLMLVENVGHMTNPAIILDRDQNEVPBG 360

QY 369 IMDALFTGLIAHGLKASDV--NGPLINGRTSGIYIVKPKMHGPAEVAFTCELFSEVEDV 426
Db 361 IMDAULTALIALH-----DVGPEGREANSRAGSMYVVKPKMHGPEEVAFAVEIFDVEAL 415

QY 427 LGLPONTMKIGIMDEBERRITVNLKACIKAAARVVFINTGFLDRTGDEIHTSMEAGPMVR 486
Db 416 LGVFRNTIKGIMDEBERRITVNLKKAIRAAERVVVFINTGFLDRTGDEIHTSMEAGPMVR 475

QY 487 KGTMSQPWILAYEDHNVDAGLAAGPSGRAQVGKGMWMTMELMADVMVETKIAQPRAGAST 546
Db 476 KGMKQAAWISAYEANNVDTGTEGGLAGHAQIGKGMWAMPDLMAAMLEQKIAHPKAGANT 535

QY 547 AWVPSPTAATLHALHVQVDVAVQOGLAGKRRATIEQLTTIPLAKELAWADEIREVED 606
Db 536 AWVPSPTAATLHALHVQVDVAVQOGLAGKRRATIEQLTTIPLAKELAWADEIREVED 595

QY 607 NNCQSILGVYVRWVDGVCCKVPDIHDVVALMEDRATLRISOLLANLWLRHGVTSADVR 666
Db 596 NNAQGIIGVYVRWVDGVCCKVPDIHDVVALMEDRATLRISOLLANLWLRHGVTSADVR 655

QY 667 ASLERMAPLVDRQAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQNGYTPILHRRRR 726
Db 656 DSLQRMALVDRQAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQNGYTPILHRRRR 715

QY 727 EFXAR 731
Db 716 ELXAR 720

RESULT 9
MASZ_AGRTS STANDARD; PRT; 731 AA.
ID MASZ_AGRT5
AC Q8UJ85;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GlsB OR ATU0047 OR AGR_C_78
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.P. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley P., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58".
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,
RA Quorlo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gordon J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58".
RL Science 294:2323-2328(2001).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.

```

CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AEO08979; AAL41078.1; ALT INIT.
CC EMBL; AEO07947; AAK85871.1; ALT_INIT.
CC PIR; AH2582; AH2582.
CC PIR; F97364; F97364.
CC HAMAP; MF_00641; -, 1.
CC InterPro; IPR001465; Malate_synthase.
CC InterPro; IPR006253; Malate_synthG.
CC Pfam; PF01274; Malate_synthase; 1.
CC TIGRFAMs; TIGR01345; malate_syn_G; 1.
CC Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
CC Complete proteome. 346
CC ACT_SITE 346 CATALYTIC BASE (BY SIMILARITY).
CC FT ACT_SITE 637 CATALYTIC ACID (BY SIMILARITY).
CC SQ SEQUENCE 731 AA; 79520 MW; 69F304D5D6F8E8FFB CRC64;
CC
CC Query Match 39.3%; Score 2380; DB 1; Length 731;
CC Best Local Similarity 63.5%; Pred. No. 1.2e-125;
CC Matches 461; Conservative 91; Mismatches 170; Indels 4; Gaps 2;
CC
CC QY 8 HTRDSVGNLRARVLYDFVNNALPGTDIDPDPSFWAGVDKVVADLTPOQALLNARDEL 67
CC Db 9 HVSRTDKFGLSIDRLRYAELTDEVLPGTGLDSETFEGPSAIVHSLSPKRELLAKRDL 68
CC
CC QY 68 QRAQDKWHRVRIEIPDMAYQFTEIGYLAPEPDDFTITTSQVDARITTAGPOLVVP 127
CC Db 69 QEKIDGWYKENG-APGDFDAYFAFLKEIGYLPPEGFGFKVETNNVDPELVAVAGPOLVVP 127
CC
CC QY 128 VLNAREPALNAARWGSGLYDALYGTIDVPIETDGAEGKPTYNKVRGDKVIAYARKEFLDSV 187
CC Db 128 VMKARVALNAARWGSGLYDALYGTIDVPIETDGAEGKPTYNKVRGDKVIAYARKEFLDSV 187
CC
CC QY 188 PLSSGSGFATGETVDQGLVVALPKSTGLANPGQFAGYTGAAESPTSVLLINGLHIE 247
CC Db 188 PLETGWSVDVTGFIADAGLQLOAIGAAATGLDKDAVQFKGFGSAAKAPATILIGKNGLTE 247
CC
CC QY 248 ILTDPESQVGTDRAGVKDVLIESATITTMDFEDSVAADAAKVLGYRNMGLNKGDLA 307
CC Db 248 IVIDPSTEIGKSDRAGISDVILESATITTMDCEDSVAADAEKVLVYGNWGLMKGDLT 307
CC
CC QY 308 AAVDKDGTAFRLVNLDRNYTPAGGQFTLPGRSLMFRVNVGHMTNDAIVDTDGSVEFE 367
CC Db 308 EAVSKGSGNTFTRLNPDRTYTPAGSALTLPGRSLMFRVNVGHMTNDAIVDTDGRDVE 367
CC
CC QY 368 GIMDALFTGLIAGHGKASDVNGPLNSRTGSIYTVKPKWHGPAEVAFTCELFSSVEDVL 427
CC Db 368 GIMDAVVTALIALYDVGPS---GRRQNSRAGSGVYVVKPKHGGPEEVAFAFENLFAVENLV 424
CC
CC QY 428 GLPQNTMKIGIMDEERTTNVTKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRK 487
CC Db 425 GNAPTNMGKIMDEERTTNVTKESITRAAKDRVVFINTGFLDRTGDEIHTSMEAGPMIRK 484
CC
CC QY 488 GTMKSQPIWILALHYHQVDVAAVQCGLAGKRATTIEQLTITPLAKELAWAPDEIRIEVDN 547
CC Db 485 GDNKQAAWIAAYENWNKVDVIGLECGSGHAGIIGKWWAMPDMLMALEQKIAHPKAGANTA 544
CC
CC QY 548 WYPSPTAATLHALHYHQVDVAAVQCGLAGKRATTIEQLTITPLAKELAWAPDEIRIEVDN 607
CC Db 545 WYPSPTAATLHALHYHQVDVAAVQCGLAGKRATTIEQLTITPLAKELAWAPDEIRIEVDN 604
CC
CC QY 608 NQSGILGYVVRWVDQGVGCKVPDHDVALMEDRATLRISQALLANLWHLHGVIITADVRA 667

```

```

Db 605 NAOGILGYVVRWVDQGVGCKVPDINNIGLMEDRATLRISAQHWANLWLRHGVVTEAQIIK 664
QY 668 SLERMAPLYDRQAGVDVATPMAPNFDSDSTAFIAAQELILSGAQQNGVTEPILHRRRE 727
Db 665 TMKREMAVVDTONAGDFPAYLPMASDFDGSVAQAAVELVLKGRQENGYTEPVLHRRLE 724
QY 728 FKARAA 733
Db 725 LKAKQA 730
RESULT 10
MASZ RHILV STANDARD; PRT; 723 AA.
AC Q937W7;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MASG.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales.
CC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
RN NCBI_Taxid=387;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF39;
RA Garcia de los Santos A., Hynes M.F.;
RT "Malate synthase gene from Rhizobium leguminosarum";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC COA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AY059637; AAL17965.1; -.
CC HAMAP; MF_00641; -, 1.
CC InterPro; IPR001465; Malate synthase.
CC InterPro; IPR006253; Malate_synthG.
CC Pfam; PF01274; Malate_synthase; 1.
CC TIGRFAMs; TIGR01345; malate_syn_G; 1.
CC Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
CC ACT_SITE 338 CATALYTIC BASE (BY SIMILARITY).
CC FT ACT_SITE 629 629 CATALYTIC ACID (BY SIMILARITY).
CC SQ SEQUENCE 723 AA; 79677 MW; 4E879906CFD64444 CRC64;
CC
CC Query Match 39.1%; Score 2367.5; DB 1; Length 723;
CC Best Local Similarity 64.5%; Pred. No. 5.8e-125;
CC Matches 464; Conservative 92; Mismatches 173; Indels 5; Gaps 3;
CC
CC QY 11 RVSYGNLRARVLYDFVNNALPGTDIDPDPSFWAGVDKVVADLTPOQALLNARDELQAQ 70
CC Db 3 RVDKNGLALETVLHDFVLEVLPGLAVDADKFPADFSAIVHDLAPKNCALLAKRDELQVK 62
CC
CC QY 71 IDKWHRRVRIEIPDMAYQFTEIGYLAPEPDDFTITTSQVDARITTAGPOLVVPVLN 130
CC Db 63 IDWYRRHG-APADMDEYQSFLEIGYLPPEGSDFFQVSTQNVDPETASAGPOLVVPVWN 121
CC
CC QY 131 ARPALNAANARWGSGLYDALYGTIDVPIETDGAEGKPTYNKVRGDKVIAYARKEFLDSVPLS 190
CC Db 122 ARVALNAANARWGSGLYDALYGTIDVPIETDGAEGKPTYNKVRGDKVIAYARKEFLDSVPLS 181
CC
CC QY 191 SGSGFDATGTVQDQGLVVALPKSTGLANPGQ- FAGYTGAESPTSVLLINHLHIEIL 249

```

```

Db 182 DCRKXDVSGFAVKDGLVVRSDIDGEOAMLTGKHFGAGYRGDAAPATHILLKNNGIHIEIV 241
Qy 250 IDPESQVGTTRAGVKDVLISAITTINDFSDVAADAKVLYGYNWGLNKGDLAAA 309
Db 242 IDAATTIGKASAHISDVWLESAITTINDCEDSAADDAEDKVVVYRNWGLMGDLQEE 301
Qy 310 VDKDGTAFRLVLRNDRNTAPGGGFTLPGRLSMFVRNVGHMTNDATVDTDGSEVEPEGI 369
Db 302 VAKGTSIRTLNPLQVAGPDGAFAFVHRSLMLVRNVGHMTNPNALDRDGEVPEGI 361
Qy 370 MDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMHGPAEVAFTCELSFRVEDVLGL 429
Db 362 MDAATITGLIALYDIGPS---GRKXNSRTGSMVWVKPKMHGPEEVAFAVEIFSRVEDALGL 418
Qy 430 PONTWKIGIMDEERTTNLAKACIKAAADRVVFINTGFLDRTGDEIHTSMEAGPMVRKGT 489
Db 419 PRNTIKGIMDEERTTNLKECIAARERVVINTGFLDRTGDEIHTSMEAGPMVRKGD 478
Qy 490 MKSQWILAYEDHNVDAGLAAGFSGRAQVGKGMWMTMELMADVMETKIAQPRAGASTAW 549
Db 479 MEQAAMISAYENWNVDIGLEGLAGHAQIGKGMWMPDLVMAMLEQKIAHPKAGANTAW 538
Qy 550 PSPTAATHALHYHOVDVAANVOQGLAGKRRATITQLLTIPLAKELAWAPDEIREVNNC 609
Db 539 PSPTAATLHATHYHVRVAVVQGLKDRARAKLSILSVVAVRPNWTPPEIQRELDNA 598
Qy 610 QSILGYVVRVWDQGVGCSKVPDIHDVALMEDRATLRISOLLANWLRHGVITTSADVRASL 669
Db 599 QGILGYVVRVWDQGVGCSKVPDINNVLGMEBRATLRISAQMANWLRHGVITTSADVRASL 659
Qy 670 ERMAPLVRONAGDVAYRPMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRREFK 729
Db 659 RMAAVVDRONASDPAYRPMAGNFDSDIAFOALDLVLKGRQPNGYTEPVLHRRRLELK 718
Qy 730 ABAA 733
Db 719 AKQA 722

RESULT 11
MASZ_RHIME
ID_MASZ_RHIME STANDARD; PRT; 723 AA.
AC Q92TA4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR R00062 OR SMC02581.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Guzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RA "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. G1CB subfamily.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
DR EMBL; AL591782; CAC41449.1; -.
DR HAMAP; MF_006441; -.
DR InterPro; IPR001465; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn G; 1.
KW Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 338 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 629 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 723 AA; 78853 MW; A0E95E8A5164BB58 CRC64;

Query Match 38.8%; Score 2351.5; DB 1; Length 723;
Best Local Similarity 62.9%; Pred. No. 4 6e-124;
Matches 455; Conservative 94; Mismatches 169; Indels 5; Gaps 3;

Qy 10 DRVSVGNLRIARVLYDFVNEALPGTDDIDPSFWAGVKVADLTPTQNALNARDELQA 69
Db 2 DRVEKYGLQIDAGLHRLVLEEAMPGTGVADRRFFSAFSDLVHDLGPKNRALLVKRDELQA 61
Qy 70 QIDKWHRRVIEPIDMDAYRQELTEITGYLLPDPDDTITTSVDVAEITTTAGQOLVVVPL 129
Db 62 RLDGWTREHG-APVDMEAYEAFREIGYLLPGPDYFVSTANVDSIATAGQOLVVVPM 120
Qy 130 NARFALNAANARWGSLYDALYGTDIVPETDGAERKGTYNKVRGDKVIAYARFLLDSVPL 189
Db 121 NARYALNAANARWGSLYDALYGTDAIETDGAERKGTYNKVRGDKVIAYARFLLDSVPL 180
Qy 190 SSGSFGDAGFTVODQOLVVALPD-KSTGLNPGQFAGYTGAAESTSVLLNHGLHIEI 248
Db 181 AAGRWSDAKSFVEGATLTTLADGTSAPRNSVQFAGYDPAAPSEIVLRNGLHIVI 240
Qy 249 LIDPESQVGTTRAGVKDVLISAITTINDFSDVAADAKVLYGYNWGLNKGDLAA 308
Db 241 VLDATTPIGKADAAGISDVVLESATITINDCEDSIAAADAEDKVVVYRNWGLMGDLSE 300
Qy 309 AVDKDGTAFRLVLRNDRNTAPGGGFTLPGRLSMFVRNVGHMTNDATVDTDGSEVFE 368
Db 301 EVTKGGRAFTRRLNPDRAVTPADGATLTLPGRLSMVLRNVGHMTNPAVLDRDGEVPE 360
Qy 369 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMHGPAEVAFTCELSFRVEDVLG 428
Db 361 LMDAMVTALIALHDIGR---NGRRANSRSGSMIVVAPKMHGPEEVAFACEIFARVAAALG 417
Qy 429 LPQNTWKIGIMDEERTTNLAKACIKAAADRVVFINTGFLDRTGDEIHTSMEAGPMVRK 488
Db 418 LPANAMKMGIMDEERTTNLKECIAARERVVINTGFLDRTGDEIHTSMEAGPMVRK 477
Qy 489 TMKSQWILAYEDHNVDAGLAAGFSGRAQVGKGMWMTMELMADVMETKIAQPRAGASTAW 548
Db 478 DMKQAPWISAYENWNVDIGLEGLAGHAQIGKGMWMPDLVMAMLEQKIVHPKAGANTAW 537
Qy 549 VPSPTAATLHATHYHOVDVAANVOQGLAGKRRATITQLLTIPLAKELAWAPDEIREVNN 608
Db 538 VPSPTAATLHATHYHVRVAVVQGLKDRARAKLSILSVVAVRPNWTPPEIQRELDNN 597
Qy 609 QCSILGYVVRVWDQGVGCSKVPDIHDVALMEDRATLRISOLLANWLRHGVITTSADVRAS 668
Db 598 AQGILGYVVRVWDQGVGCSKVPDINNVLGMEBRATLRISAQMANWLRHGVITTSADVRAS 657
Qy 669 LESMAPLVRONAGDVAYRPMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRREF 728
Db 658 RMAAVVDRONASDPAYRPMAGNFDSDIAFOALDLVLKGRQPNGYTEPVLHRRRLEL 717
Qy 729 KAR 731
-----

```



```

Db      718 KAK 720
RESULT 12
MASZ BRUWE
ID MASZ BRUWE STANDARD; PRT; 728 AA.
AC Q8YIR3;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GUCB OR BME10380.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20030109; PubMed=11756688;
RA DelVecchio V.G., Kapural V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Hasekorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
-----
DR DR EMBL; AE009480; AAL51561.1; -
DR DR PIR; AF3299; AF3299.
DR DR HAMAP; ME_00641; -; 1.
DR DR InterPro; IPR001465; Malate_synthase.
DR DR Pfam; PF01274; Malate_synthase; 1.
DR DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
DR DR Trnsterase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
PT ACT_SITE 345 345 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 636 636 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 728 AA; 79985 MW; E1313A617979270D CRC64;
Query Match 38.2%; Score 23.18; DB 1; Length 728;
Best Local Similarity 61.8%; Pred. No. 3.4e-122;
Matches 444; Conservative 97; Mismatches 174; Indels 4; Gaps 2;
QY 12 VSVGNRIARVLVDFVNNALPOTDIDPDSFAGVDKVVADITPQNALNNARDELQAI 71
Db 10 VEIEGLAVPELVEFLAKEAAGTGPPEKFWKFAAIIIDLPKRNALLAKRDELQARI 69
QY 72 DKHRRRVIEPIDMDVROFLTEIGVLLPEPDDFTITSGVDABITTTAGPOLVVPVLA 131
Db 70 DANYKENRDKGYQADYQQFLKDIGVLLPEGGAFSVSTTNVPEITHIAGPOLVVPVMA 129
QY 132 RFALNANARWGLXDALYGVTDVPTDGAEGKFTYKVRGDKVIAYAKFIDDSVPLSS 191
Db 130 RYALNANARWGLYDALYGTDAISADGAERKGYKGNPKRGEKVIAWAKNFDESAPLST 189
QY 192 GSFGDATGFTVQDGLVVALPDKS-TGLANPGQFAGYTGAAESPTSVLLINHLHIEILI 250

```

```

Db      190 GKWADVAGLA VNDGKLEIRLTGDSATTLKDESQFKGYNGDAASPTNVLIAKHNHVDIVI 249
QY 251 DPESQVGTDRAGVKOVILESAITTTIMDFEDSVAADDAADKVLGYRNWGLNGDLAAAV 310
Db 250 NADHPICKTDPAHIAADVLESAISTIQDCEDSAAVDAEDKVAAYRNWGLNGKLEDTF 309
QY 311 DKGTAFRLVLRNDRNYTAPGGQFTPLGRSLMFRVNVGHMTNDIAVPTDGGSEVEFEGIM 370
Db 310 EKNGKQMTRELNGDRTYTAPDGTSTLLKGRSLMFRVNVGHMTNPAILDAEGNEVPEGIM 369
QY 371 DALFTGLIAHGLKASDVNGPLNSRTGSIYVVKPMHGFPAVFTCELPSEVEDVLGLP 430
Db 370 DAAFTSLIALHDI---GPNGRHMNSREGSVIYVVKPMHGFPAVFAFANEIPTRTTEMLGMK 426
QY 431 QNTMKIGIMDEERTTNNKAKICAAADRNVFINTGFLDRTGDEIHTSMEAGPMVRKGTM 490
Db 427 PNTLKIGIMDEERTTNNKAKICAAADRNVFINTGFLDRTGDEIHTSMEAGPMVRKGTM 486
QY 491 KSPWILAYEDHNVNDAAGLAAAGPGRAGVQKGMVTMTLMVADVETKIAQPRAGASTAVP 550
Db 487 KQAAWIGAYEQWNVDIQLEGLSCHAQIGKGMWAMPDMVAAMLEQKIAHPKAGANTAVP 546
QY 551 SPTAATLHALHYHOVDVAAVQOGLAGKRRATIEOLLTIPLAKELAWAPDEIREVDNNCQ 610
Db 547 SPTAATLHALHYKIDVAAVQEKLSRPRAKLDDILSVPAVVPNTPTDDIOHEIDNNAQ 606
QY 611 SILGYVVRVVDQGVGSKVPDIHDVALMEDRATIRISSQLLANLWLRHGVITSADVASLE 670
Db 607 GILGYVVRVVDQGVGSKVPDIHDVALMEDRATIRISSQLLANLWLRHGVITSADVASLE 666
QY 671 RMAPLVDRQAGDVAVPEMAPNPDSDTAFLLAAGLILSGAQQNGYTERPILHRRRREF 729
Db 667 RMAIVDKQNGEPLXRPMAADFCKSIAPQACDLVFKGREQNGYTERPILHRRRLELK 725
RESULT 13
MASZ BRUSU
ID MASZ BRUSU STANDARD; PRT; 728 AA.
AC Q8FZ50;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR BR1648.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

-----
DR EMBL; AE014458; AAN30550.1; --
DR TIGR; BR1648; --
DR HAVAP; MF 00641; --
DR InterPro; IPR001465; Malate_synthase.
DR InterPro; IPR006253; Malate_synthG.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn G; 1.
KW Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 345 345 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 636 636 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 728 AA; 79966 MW; F95669D002A14EDE CRC64;

Query Match      38.2%; Score 2313; DB 1; Length 728;
Best Local Similarity 61.6%; Pred. No. 6.6e-122;
Matches 443; Conservative 97; Mismatches 175; Indels 4; Gaps 2;

Qy 12 VSVGNLRIRARVLYDFVNNALPGTDIDPDSTWAGVDKVVADLTQNCALLNARDELQAOI 71
Db 10 VEIEGLAVAPELVEFLAKEAPGTGVEPEKFWGFAAIIIRDLAPKNRALLAKRDELQARI 69
Qy 72 DKWRRRRIEIPIDMDAYRQFTEIGYLLPEPDDFTITTSVGYDAEITTTAGPQLVVPVINA 131
Db 70 DAWKENRDKYSQADYQFELKIDGLYLPGEAGSVSTINVDPEITHIAGPQLVVPVINA 129
Qy 132 RFALNAANARNGSLYDALYGTVDVPEITDGAEGKPTYNKVRGDKVIAYARKFLDSDVPLSS 191
Db 130 RVALNAANARNGSLYDALYGTDAISEADGAEGKGYNPKRGEKVIAMAKNFDLSDSAPLST 189
Qy 192 GSFGDATFTVDGQGLVALPKDS--TGLANPQFAGYTCAESPTSVLLINHGHIIEILI 250
Db 190 GKWADVAGLVNDGKLEIRLTGDSATTLKDSQFKYNGDAASPTNVLAKHNHVDIVI 249
Qy 251 DPESVGTTDRAGVKVDVILESAITTIMDFEDSVAADAAKVLGVNRNWLGNKGDLAAYV 310
Db 250 NADHPIGKTDPAHIADVLESASITIQCEDSIAAADAEDKAVYVRNWLGLNKLJEDTF 309
Qy 311 KDKGTAFURVLRNDRNYTAPGGQTLPLGRSLMFVRNVNHLMTNDAIVDDTDCSEVFEGIM 370
Db 310 EKNGKQMTRLNGDRITYTAPDGSTLTLLKSHSLMLVRNVNHLMTNPAILDAEGNEVPVEGIM 369
Qy 371 DALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMGHGAEPVAFCTCLFSRVEDVLGLP 430
Db 370 DAAFTSLIALHDI----GNGRHNRSGSVIIVKPKMGHGEVAFANEIFRTEEMLGK 426
Qy 431 QNTMKIGIMDEERTTVNLKACIKAAADRVVVFINTGFLDRTGDEIHTSMEAGPVRKGTM 490
Db 427 PNTLKIGIMDEERTTVNLKAZAIRAAKDRVVVFINTGFLDRTGDEIHTSMEAGPMIRKGM 486
Qy 491 KSPWILAYEDHNDVADGLAAGSGPAQVKGKWTMTLMADVETKIAQPRAGASTAWVP 550
Db 487 KQAAWIGAYEQNVNDIGLECGLSHAQIKGKNWAMPDYMMAALFQKIAHPKAGANTAWVP 546
Qy 551 SPTAATLHALHVQDVAAVQOGLAGKERATEQLTTIPLAKELAWAPDEIREEVNQCQ 610
Db 547 SPTAATLHAHVHKIDVAQVEKLKSRPRAKLDDILSVFVAVRPNWTPDDIQHIDNNAQ 606
Qy 611 SILGVVVRWVGCGSKVPDIHDVVALMEDRATLRISSQLLANLHRHVITTSADVRSLE 670
Db 607 GLGVVVRWIDVGCGSKVPDINNVLGMLMEDRATLRISSAQHLANWLYHGVSQAQVMTMK 666
Qy 671 RNAPVLVDRONAGDVAVRMAPNFDSDIAFLAAQELILSGAOPNGYETPIILHRRRREP 729
Db 667 RNAAIVDKQNEGDPLVYRPMADFDKSIQAFQAACDLVFKRGQPNGYETPVILHRRLELK 725

STANDARD; PRT; 738 AA.
ID _MASZ_CORGL

```

RESULT 14
 MASZ_CORGL

```

AC P42450;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR ACB OR CGL2329.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-10, AND CHARACTERIZATION.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RX MEDLINE=95111631; PubMed=7812449;
RA Reinscheid D.J., Eikmanns B.J., Sahm H.;
RT "Malate synthase from Corynebacterium glutamicum: sequence analysis
  of the gene and biochemical characterization of the enzyme.";
RL Microbiology 140:3099-3108(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13059 / AS019;
RA Lee H.S., Sinskey A.J.;
RT "Molecular characterization of aceB, a gene encoding malate synthase
  in Corynebacterium glutamicum.";
RL J. Microbiol. Biotechnol. 4:256-263(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
  CoA.
CC -1- ENZYME REGULATION: Inhibited by oxalate, glycolate and ATP.
CC -1- PATHWAY: Glyoxylate bypass; second step.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the malate synthase family. GLCB subfamily.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
```

EMBL; X78491; CAA55243.1; --

EMBL; L27123; AAA68074.1; --

EMBL; AP005281; BAB99722.1; --

PIR; I40715; I40715.

HSP; P37330; I08C.

HMAP; MF 00641; -- 1.

InterPro; IPR001465; Malate_synthase.

InterPro; IPR006253; Malate_synthG.

Pfam; PF01274; Malate_synthase; 1.

TIGRFAMs; TIGR01345; malate syn G; 1.

Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;

Complete proteome.

INIT_MET 0 0

ACT_SITE 355 355 CATALYTIC BASE (BY SIMILARITY).

ACT_SITE 646 646 CATALYTIC ACID (BY SIMILARITY).

SEQUENCE 738 AA; 82231 MW; 72AA0663AE7C87F4 CRC64;

Query Match 37.8%; Score 2289; DB 1; Length 738;

Best Local Similarity 61.3%; Pred. No. 1.5e-120;

Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;

Qy 9 TDRVSVNLRIRARVLYDFVNNALPGTDIDPDSTWAGVDKVVADLTQNCALLNARDELQ 68

Db 19 TERVDAGGMQVAKVLYDFVTEAVLPRVGVDAERKFSGFAAIARDLTPRNRELLARDELQ 78

Qy 69 AQIDKWHRRRRIEIPIDMDAYRQFTEIGYLLPEPDDFTITTSVGYDAEITTTAGPQLVVPV 128

Db 79 MLIDDDVHRNN-STIQEAYEDFKIGLVVEEPEAAERTQVNDTEISTAGPQLWPEI 137
 Qy 129 LNARFALNANARWGSYDALYGTVDIPETDGAEGKFTYKVRGDKVIAAKFDDSDVP 188
 Db 138 LNARFALNANARWGSYDALYGTNALPETDGAEGKFTYKVRGDKVIAAKFDDSDVP 197
 Qy 189 LSSGSGDATGFTVQDQQLVVALPKSTGLANPQAGTGAESTSVLLINHLGHIEI 248
 Db 198 LDCASHADVEKNITDKLAHIGDSYRLKNSRYGFTGNFLDPEAILLETNGLHIEL 257
 Qy 249 LIDPESQVGTTRDRAGVKVDVLESATITIMDFEDSVAADAAKVLGYRNWGLNGKDLAA 308
 Db 258 QIDPVPHIGKADTKGLKDIVLESATITIMDFEDSVAADAAKVLGYRNWGLNGKDLAA 317
 Qy 309 AVKDGTAFLVRLNRDNYTAPGGQFTLPGRSLMFRVNVGHLMNTDAIVDTGSEVFEG 368
 Db 318 EMKNGRIPTRELKORVITGRNGTELVLHGRSLLFVRNVGHLMQNPSTL-IDGEEIPEG 376
 Qy 369 IMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLG 428
 Db 377 IMDAVITTCALPGIAPQNN--KORNSRKGSIYIVKPKMGPAEVAFTCELSRVEDVLG 433
 Qy 429 LPQNTMKIGIMDEERTTNVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMAGPMVRKG 488
 Db 434 LPRHTLVGVNDEERTTSVNLDAIMEVADRLAFINTGLDRTGDEIHTSMAGPMVRKA 493
 Qy 489 TMSQWILAYEDHNVDAAGLAAGFSRAQVKGKMTWMTLMADMTETKIAQPRAGASTAW 548
 Db 494 DMQAPWKQAYEDNNVDAGIQORGLPKAQKGMWMTLMADMTETKIAQPRAGASTAW 553
 Qy 549 VPSPTAATLHALHYQVDVAAVQOGL-AGKRRTATIEQLTIPLAKELAWAPDIREVDN 607
 Db 554 VPSPTGATLHATHYRVDVFKVDELRAAGRDSLRNLTITPANTNWSBEKKEMDN 613
 Qy 608 NCOSILGVYVVRWVQGVGSKVPDIHDVLMEDRATLRISQQLANLWLRHGVITSADVRA 667
 Db 614 NCOSILGVYVVRWVQGVGSKVPDIHDVLMEDRATLRISQQLANLWLRHGVITSADVRA 673
 Qy 668 SLERMAPLVDQRQAGVAYEPMAFNPDSDTAFIAQELILSGAQQNGYTEPILHRRRE 727
 Db 674 SLERMAVVDQKQAGDEATVAPNDASLAFQAAKDLIFEGTSPSGITEPILHRRRE 733
 Qy 728 FKAR 731
 Db 734 FKAK 737

RESULT 15
 MASZ_COREF STANDARD; PRT; 748 AA.
 AC Q8FNE3;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Malate synthase G (EC 2.3.3.9).
 GN GUCB OR MASZ OR CR2231.
 OS Corynebacterium efficiens.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=152794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
 RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
 RA Usuda Y., Sugimoto S.,
 RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
 CoA.
 CC -!- PATHWAY: Glyoxylate bypass; second step.
 CC -!- SUBUNIT: Monomer (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AP005221; BAC19041.1; -
 DR HAVAP: MF_006411; -; 1
 DR InterPro: IPR001465; Malate synthase.
 DR InterPro: IPR006253; Malate synthase.
 DR Pfam: PF01274; Malate synthase; 1.
 DR TIGRFAMs: TIGR01345; malate syn G; 1.
 KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
 KW Complete proteome.
 FT ACT_SITE 362 362 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 653 653 CATALYTIC ACID (BY SIMILARITY).
 SQ SEQUENCE 748 AA; 83491 MW; F9550473EC4E9A09 CRC64;
 Query Match 37.5%; Score 2273; DB 1; Length 748;
 Best Local Similarity 61.5%; Pred. No. 1.2e-119;
 Matches 444; Conservative 90; Mismatches 182; Indels 6; Gaps 4;
 Qy 9 TDRVSVGNLRIARVLYDFVNEALPGTIDIPDPSFWAGVKVADLTPOQALLNARDELQ 68
 Db 26 TERYTVGGMQVAKVLRDLFTESVLPVRGVDAERFWNGFGDIVRDMTPRNLAREDELQ 85
 Qy 69 ACDKWHRRVRIEPLDMAYRQFLTEIGYLLPPDPDTITTSVDALITTTAGPQLWVPV 128
 Db 86 AOLDEYRENPGR-DPEKYEAFLEIGYLVDPAPAEITQNDISEIATTAGPQLWVPV 144
 Qy 129 LNARFALNANARWGSYDALYGTVDIPETDGAEGKFTYKVRGDKVIAAKFDDSDVP 188
 Db 145 LNARFALNANARWGSYDALYGTNAIPDEDGAERGAENPVRGQKVIQNGRDFDLAVLP 204
 Qy 189 LSSGSGDATGFTVQDQQLVVALPKSTGLANPQAGTGAESTSVLLINHLGHIEI 248
 Db 205 LDCASHADVEKNITDKLAHIGDSYRLKNSRYGFTGNFLDPEAILLETNGLHIEL 264
 Qy 249 LIDPESQVGTTRDRAGVKVDVLESATITIMDFEDSVAADAAKVLGYRNWGLNGKDLAA 308
 Db 265 QIDPVPHIGKADTKGLKDIVLESATITIMDFEDSVAADAAKVLGYRNWGLNGKDLAA 324
 Qy 309 AVKDGTAFLVRLNRDNYTAPGGQFTLPGRSLMFRVNVGHLMNTDAIVDTGSEVFEG 368
 Db 325 EVAKGDEFTFRKNDNRVFGKNGAELTLGRSLLFVRNVGHLMNTNPAIL-VDGEIYEG 383
 Qy 369 IMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLG 428
 Db 384 IMDAITTCALPGIAPQNN--NSRKGSIYIVKPKMGPAEVAFTCELSRVEDVLG 440
 Qy 429 LPQNTMKIGIMDEERTTNVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMAGPMVRKG 488
 Db 441 LPRHTLVGVNDEERTTSVNLDAIMEVADRLAFINTGLDRTGDEIHTSMAGPMVRKA 500
 Qy 489 TMSQWILAYEDHNVDAAGLAAGFSRAQVKGKMTWMTLMADMTETKIAQPRAGASTAW 548
 Db 501 DMQAPWKQAYEDNNVDAGIQORGLPKAQKGMWMTLMADMTETKIAQPRAGASTAW 560
 Qy 549 VPSPTAATLHALHYQVDVAAVQOGL-AGKRRTATIEQLTIPLAKELAWAPDIREVDN 607
 Db 561 VPSPTGATLHATHYRVDVFKVDELRAAGRDSLRNLTITPANTNWSBEKKEMDN 620
 Qy 608 NCOSILGVYVVRWVQGVGSKVPDIHDVLMEDRATLRISQQLANLWLRHGVITSADVRA 667
 Db 621 NCOSILGVYVVRWVQGVGSKVPDIHDVLMEDRATLRISQQLANLWLRHGVITSADVRA 680
 Qy 668 SLERMAPLVDQRQAGVAYEPMAFNPDSDTAFIAQELILSGAQQNGYTEPILHRRRE 727

Db 681 SLERMAVVVEQNAQDPNYLNNAPNFTESVAFOAARDLILKGTESPAGYTEPILHARRR 740
QY 728 FX 729
Db 741 FX 742

Search completed: November 21, 2003, 16:04:22
Job time : 15.8195 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 57.6698 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-52
Perfect score: 6061
Sequence: 1 MQHHHHTDRVSGNLRIA.....SAATRRPCTGRDGRMACQ 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2634	43.5	724	2 Q9AE55	Q9ae55 rhodococcus
2	2569	42.4	725	16 Q9I636	Q9i636 pseudomonas
3	2416	39.9	721	16 Q98DK4	Q98dk4 rhizobium l
4	2380	39.3	744	16 Q9CJ85	Q9cjb85 agrobacteri
5	2367.5	39.1	723	2 Q937W7	Q937w7 rhizobium l
6	2351.5	38.8	723	16 Q92TA4	Q92ta4 rhizobium m
7	2318	38.2	728	16 Q9YIR3	Q9yir3 brucella me
8	2313	38.2	728	16 Q9FZ50	Q9fz50 brucella su
9	2273	37.5	748	16 Q9FN83	Q9fnb3 corynebacte
10	2258	37.3	727	16 Q9KB03	Q9kb03 bacillus ha
11	2168.5	35.8	723	16 Q9FDN6	Q9fdn6 escherichia
12	1545.5	25.5	421	16 Q9CB77	Q9cb77 mycobacteri
13	1272.5	21.0	421	2 Q8KQ29	Q8kq29 streptomyce
14	1267.5	20.9	421	2 Q9CQ25	Q9cq25 amycolatops
15	1232.5	20.3	421	2 Q93C57	Q93c57 corynebacte
16	1228.5	20.3	421	2 Q9AE85	Q9ae85 nocardia la

Q93C55 corynebacte
Q9XA17 streptomyce
Q8GIQ5 streptomyce
Q93TX4 stigmatella
P74569 synchocyst
Q8F865 leptospira
Q8RU02 methylobact
Q8DHW0 synchococc
Q8YR14 anabaena sp
Q8G6Y6 bifidobacte
Q9JTN3 neisseria m
Q9JYN6 neisseria m
Q9K8B0 bacillus ha
Q9PHT4 campylobact
Q9WZ17 thermotoga
Q9XIK6 thermotoga
Q8Y073 raistonia s
Q92CH4 listeria in
Q9A9W8 caulobacter
Q8YJN9 listeria mo
Q8YJ97 brucella me
Q8FYK1 brucella su
Q9EV19 bradyrhizob
Q98G91 rhizobium l
Q8UC0 agrobacteri
Q92MK3 rhizobium m
Q9KUH4 vibrio chol
Q8EBS2 shewanella
Q8RA48 thermoanaer

ALIGNMENTS

RESULT 1

Q9AE55 PRELIMINARY; PRT; 724 AA.
AC Q9AE55;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Malate synthase.
GN VICA.
OS Rhodococcus fascians.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1828;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D188;
RA Vereecke D.M., Cornelis K., Van Montagu M., El Jaziri M., Holsters M.,
RA Goethals K.;
RT "Characterization of a chromosomal locus that affects pathogenicity in
RT Rhodococcus fascians."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AJ301559; CAC35701.1; -
DR HSSP; P37330; 1D8C.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthG.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
SQ SEQUENCE 724 AA; 78609 MW; F889FE883890995E CRC64;
Query Match 43.5%; Score 2634; DB 2; Length 724;
Best Local Similarity 69.0%; Pred. No. 3.2e-144;
Matches 499; Conservative 82; Mismatches 138; Indels 4; Gaps 2;
QY 9 TDRVSVGNLRIARVLYDFVNEALPGDIDPDSFWAGVDKVVADLTQNALNARELQ 68
DB 2 TDRVQAGGLQVAKVLFDFVEKEALPGDIDPDSFWAGVAGAAVIAADLAEKALLAVRDEIQ 61
QY 69 AOIKDWRRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTTAGPOLVVPV 128

Db 62 GKVDWHEGHEAGAEYDRAAYKAFKEIGYLLDDEPAPQIHTSGVDTEITTTAGPQLVVPV 121

Qy 129 LNARFALNANARWGSYDLYGTDVTPETDGAEGKGYTNKVGDKVIAVARKFLDSDVP 188

Db 122 LNARFALNANARWGSYDLYGTDVTPETDGAEGKGYTNKVGDKVIAVARKFLDSDVP 181

Qy 189 LSSGSGDGTGFTVQDQGVVLPDKST-GIANPGOPAGYTGAEPSPTSLINLHGHIE 247

Db 182 LSSGSHVGTGYVDAASLTVTADGSTVGLKDPQSLLGYQGTPDAPTALFLFVHGHFE 241

Qy 248 ILIDPESQVGTDRAGVKVILESAITTIMDFDSVAADAAVKVLYGRNWLGLNKGDLA 307

Db 242 IQIDPESPIGKTGAGVKVILESAITTIMDFDSVAADAAVKVLYGRNWLGLNKGDLT 301

Qy 308 AAVDKGTAFILVLRNDRNTAPGGGFTLPGSLMFRVNVGHMTNDALVDTDGSEVFE 367

Db 302 EESKGGKTFTRANKKORTYTSVDGSELTILHGRSLFVRNVGHMTNDALVDGNEVPE 361

Qy 368 GIMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVL 427

Db 362 GILDALFTSLAGLSLTPDNV---LNSRTGSLYIVKPKMHGPAEVAFTCELSRVEDVL 418

Qy 428 GLPONTWKIGIMDEERTTNVNLKACIKAAADRVVFNITGFLDRTGDEIHTSMEAGNVRK 487

Db 419 GLPNTLTKVIMDEERTTNVNLKACIKAAADRVVFNITGFLDRTGDEIHTSMEAGNVRK 478

Qy 488 GTMKSQPWILAYEDHNVDAGLAAGFSGRAQVKGWMTMTLMADMTETKIAQPRAGASTA 547

Db 479 GAKKGEKIAAYEDFNVDTGLAAGLQKQAIKGKMWAMPDLMLEQKIGHPKAGANTA 538

Qy 548 WVPSPATAATLHALHYHVDVAAVQOGLAGKRRATIEQLTIPLAKELAWAPDEIREVDN 607

Db 539 WVPSPATAATLHALHYHVDVAAVQOGLAGKRRATIEQLTIPLAKELAWAPDEIREVDN 598

Qy 608 NCOSILGYVVRWIDQGVGCKVDPIDHDVALMEDRATILRSSQLLANLHGHVITSADVRA 667

Db 599 NSQSGILGYVVRWIDQGVGCKVDPIDHDVALMEDRATILRSSQLLANLHGHVITSADVRA 658

Qy 668 SLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRE 727

Db 659 SLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRE 718

Qy 728 PKA 730

Db 719 YKA 721

RESULT 2

Q91636 PRELIMINARY; PRT; 725 AA.

AC Q91636

DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Malate synthase G.

GN GLCB OR PA0482

OS Pseudomonas aeruginosa

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RA MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.B., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen."

RL Nature 406:959-964 (2000).

DR EMBL; AE004485; AAG03871.1; --

DR HSSP; P37330; ID8C.

DR InterPro; IPR001465; Malate synthase.

DR InterPro; IPR006253; Malate synthase.

DR Pfam; PF01274; Malate synthase; 1.

DR TIGRfam; TIGR01345; malate_syn_G; 1.

KW Complete proteome.

QY SEQUENCE 725 AA; 78659 MW; 3669670A9E38D391 CRC64;

Query Match 42.4%; Score 2569; DB 16; Length 725;

Best Local Similarity 69.1%; Pred. No. 1.9e-140;

Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;

Qy 9 TDRVSVGNLRTARVLYDFVNNELPGTIDDPDSFWAGVGVVADLTQNCALLNARELO 68

Db 2 TERVQVGLQVAKVLFDFVNNELIPGTGVSADTWTGAEAVINDLAPKNKALLAKRDELQ 61

Qy 69 AQIDKWHRRRVIEPIDMDAYRQFLTEIGYLLPEPDFTITTTSGVDABEITTTAGQLVVPV 128

Db 62 AKIDGWHQARAGQAHDVAAYKAFLEEIGYLLPEAEDFQAGTQNVDDDEIARMAGQLVVPV 121

Qy 129 LNARFALNANARWGSYDLYGTDVTPETDGAEGKGYTNKVGDKVIAVARKFLDSDVP 188

Db 122 LNARFALNANARWGSYDLYGTDVTPETDGAEGKGYTNKVGDKVIAVARKFLDSDVP 181

Qy 189 LSSGSGDGTGFTVQDQGVVLPDKST-TGLANPGQFAGYTGAEPSPTSLINLHGHIE 247

Db 182 LSSGSHVGTGYVDAASLTVTADGSTVGLKDPQSLLGYQGTPDAPTALFLFVHGHFE 241

Qy 248 ILIDPESQVGTDRAGVKVILESAITTIMDFDSVAADAAVKVLYGRNWLGLNKGDLA 307

Db 242 IQIDPSSPVGQTDAGVKVILMEALATITIMDCESVAADADDKVITYRNWLGLMKGDLA 301

Qy 308 AAVDKGTAFILVLRNDRNTAPGGGFTLPGSLMFRVNVGHMTNDALVDTDGSEVFE 367

Db 302 EESKGGKTFTRANKKORTYTSVDGSELTILHGRSLFVRNVGHMTNDALVDGNEVPE 361

Qy 368 GIMDALFTGLIAHGLKASDVNGPLI--NSRTGSIYIVKPKMHGPAEVAFTCELSRVED 425

Db 362 GIQDGLFTSLIAH----DLNGNTSRKNGRTGSIYIVKPKMHGPEEAFTNELFGRVED 416

Qy 426 VLGI.PONTMKGIMDEERTTNVNLKACIKAAADRVVFNITGFLDRTGDEIHTSMEAGPMV 485

Db 417 VLGLPNTLTKVIMDEERTTNVNLKACIKAAADRVVFNITGFLDRTGDEIHTSMEAGAVV 476

Qy 486 RGTMKSQPWILAYEDHNVDAGLAAGFSGRAQVKGWMTMTLMADMTETKIAQPRAGAS 545

Db 477 RKGAMKSEKMGAYENNNVDVGLATGLQKQAIKGKMWAMPDLMLEQKIGHPKLAGAN 536

Qy 546 TAVVPSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLTIPLAKELAWAPDEIREEV 605

Db 537 TAVVPSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLTIPLAKELAWAPDEIREEV 596

Qy 606 DNNQCSILGYVVRWIDQGVGCKVDPIDHDVALMEDRATILRSSQLLANLHGHVITSADV 665

Db 597 DNNQCSILGYVVRWIDQGVGCKVDPIDHDVALMEDRATILRSSQLLANLHGHVITSADV 656

Qy 666 RASLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRR 725

Db 657 VESLKRMAVVVDQRNAGSDPSYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRR 716

Qy 726 REFKA 731

Db 717 REFKA 722

RESULT 3

Q98DK4

ID Q98DK4 PRELIMINARY; PRT; 721 AA.

AC Q98DK4

DT 01-OCT-2001 (TReMBLrel. 18, Created)

DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

```
DE Malate synthase G.
GN MLR4664.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFP303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AF003004; BAB51267.1; -.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
KW Complete proteome.
SQ SEQUENCE 721 AA; 78058 MW; 65376311A7E1BDFP CRC64;

Query Match 39.9%; Score 2416; DB 16; Length 721;
Best Local Similarity 64.4%; Pred. No. 1.4e-131;
Matches 467; Conservative 86; Mismatches 164; Indels 8; Gaps 3;

QY 9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
DB 2 TDRISIAGLRIAGEIHDVAGALPGTGLAADFWSGSAIVHDIAPKRALXKRDAMQ 61
QY 69 AQIDKWHRRRVIEPIDMDAYROFLTEIGVLLPEPDDFTITTSVDABITTTAGPOLVVPV 128
DB 62 ERLDQWYRDNG-APVDMVEYKSFLEIGLYVPEGPAFVSVDNDVPEIIVAVAGQLVVPV 120
QY 129 VNARFALNAANRWGSLVDALYGTVDVIVETDGAEGKPTYNKVRGDKVIYARKFLDQSV 188
DB 121 MNARYALNAANRWGSLVDALYGTDAIPEITGAEKGFNPARGAKVIAWAKDFLDQSV 180
QY 189 LSSGSFGDATGTQVQDGLWALPKSTGLANPGQFAGVTGAESPTSVLLINGLHIEI 248
DB 181 LTSQKWAGVGLSVAHGALKGACAGGTTLADPQFAGYRGDAANDPDLVVKLGLHIEI 240
QY 249 LIDPSQVGTTRAGVKDVILESAITTIMDPEDSVAADAAADKVLGYRNWGLNKGDLAA 308
DB 241 VVDRNQIGRTDPAGIADVILESAITTIQDCEDSVAADAAQDKVVVYRNWGLNKGDLAE 300
QY 309 AVDKDGTAFRLVNDERNYTPAGGQFTLPGRSLMFVNVNGLMTNDIAIVTDSEVPEG 368
DB 301 EISKAGRSFVRKLNADRAYTAPAGCQITVPGRSMLVNVNGLMTNPAILDRDQNEVPEG 360
QY 369 IMDALFTGLIATHGLKASDV--NGPLNSRTGSIYIVPKMGPAEVAFTCELSRVEDV 426
DB 361 INDAALTALIALH-----DVGPEGRANSRAGSMYVVKPMGPEEVAFAVEIDRVEAL 415
QY 427 LGLPONTWKIGIMDBERTTNLKCIIKAAADRVVFINTEGLDRTGDSIHSTMGAGMVR 486
DB 416 LGMPRTNIMGIMDBERTTNLKEAIRAARERVVFINTEGLDRTGDSIHSTMGAGPMIR 475
QY 487 KGTMSQPWILAYEDHNDYDAGLAAGSGRAQVGKGMWMTMLMDMVTETKIAQPRAGAST 546
DB 476 KDMKQAAWISAYEAWNVDTGLECGLAGHAQIGKGMWNPDMMAAMLEQKLAHPKAGANT 535
QY 547 AVWPSPTAATLHLYHQVDVAVOOGLAGKRRATIEQLLTPIAKELANAPDEIRBEVD 606
DB 536 AVWPSPTAATLHLYHQVDVAVQAALKSRPKAKLDDILSVFAVAVRNWTPDSIQRELD 595
QY 607 NNCQSLGIVYVWVDQGCCKVPDIHDVLMEDRATLRISSOLLANWLRHGVITSADYR 666
DB 596 NNAQGLIGVYVWVDQGCCKVPDINDVGLMEDRATLRISSOLLANWLRHGVITSADYR 655
```

```
QY 667 ASLERMAPLVDRONAGDVAYRVPAPNFDSDISIAFAAQELILSGAQOQNGYTEILHRRR 726
DB 656 DSQRMAALVDRONVGDPLRYRNPAPDFDKSIAFAACADLVFKGTSQPNGYTEPVLHARR 715
QY 727 EFKAR 731
DB 716 ELKAQ 720

RESULT 4
Q8UJ85
ID Q8UJ85 PRELIMINARY; PRT; 744 AA.
AC Q8UJ85;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Malate synthase G.
GN GICB OR ATU0047 OR AGR_C_78.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kuyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liac L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollio B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
RA Hummel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA William C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lenc C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL; AE008979; AAL41078.1; -.
DR EMBL; AE007947; AAK85871.1; -.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
KW Complete proteome.
SQ SEQUENCE 744 AA; 80991 MW; FB74ED358D50319 CRC64;

Query Match 39.3%; Score 2380; DB 16; Length 744;
Best Local Similarity 63.5%; Pred. No. 1.8e-129;
Matches 461; Conservative 91; Mismatches 170; Indels 4; Gaps 2;

QY 8 HTRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDEL 67
DB 22 HVSRTDKFGLSIDRLYAFALTDEVLPGTGLESETFEFSIAIVHLSPKNRELLAKRDAL 81
QY 68 QAQIDKWHRRRVIEPIDMDAYROFLTEIGVLLPEPDDFTITTSVDABITTTAGPOLVVP 127
DB 82 QEKIDGWYRENG-APGDFDIAYEALKEIGVLLPEGPGFVETNNVDPEIIVAVAGPOLVVP 140
QY 128 VNARFALNAANRWGSLVDALYGTVDVIVETDGAEGKPTYNKVRGDKVIYARKFLDQSV 187
```

```

Db 141 VMARYALNANRWGSLYDALYGTDAISDADGAERGRVNPGRGKVIAMARNFLDESA 200
Qy 188 PLSSGFGDAGTGTVDQGVQVLPALPKSGTGLANPGFAGYTGABESPTSVLLINHLGHIE 247
Db 201 PLETGMSDVTGFNIADGLLQALGAATTKGLDAVQFKGSGEAAKPAATILLGKNGLHTE 260
Qy 248 ILDPESQVGTTRDRAGVKDVILESATTTIMDFEDSVAADADKVLGYRNWLGKNGDLA 307
Db 261 IVIDPSTEIGKSRAGISDVILESATTTIMDCEDSVAADADKVLGYRNWLGKNGDLT 320
Qy 308 AAVDRKGTAFRLVLRNDRNTYAPGGQFTLPGRSLMPFVRNVGHMTNDIAIVTDGSEVFE 367
Db 321 EAVSKGCGNTFTRILNPDYATPDGSAITLPGRLSLVRNVGHMTNPAILDGRDQDVE 380
Qy 368 GIMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMGHGAFAVFTCELFPSRVEDVL 427
Db 381 GIMDAVVTAIALYDVGPS---GRRQNSRAGSIYIVKPKMGHGAFAVANEIFARVENUV 437
Qy 428 GLPQNTMKIGIMDEERTTIVNLKACIKAAADRVFINTGFLDRTGDEIHTSMEAGPMVRK 487
Db 438 GMAPNTMKGIMDEERTTIVNLKESIRAAKDRVVFINTGFLDRTGDEIHTSMEAGPMIRK 497
Qy 488 GTMKSQWILAYEDHNVDAAGLAFSGRAQVKGKMTNTELMADVETKIAQPRAGASTA 547
Db 498 GDMKQAAMIAAYENWNVDIGLECGLSGHAQIGKMWMPDLMAAMLEQKIAHPKAGANTA 557
Qy 548 WFSPTAATLHLYHGVDAVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPOBIREVDN 607
Db 558 WFSPTAATLHLYHGVDAVAAVQOGLKSGRAKLSLISVPAVPRNPTPEIQEELDN 617
Qy 608 NCQSILGYVVRWVQGVGCKVDPIDHVALMEDRATLRISSQLLANLWLRHGVITSADVRA 667
Db 618 NAQOILGYVVRWVQGVGCKVDPIDNIGLMDRATLRISSAQMANKLWLRHGVVTEAQIIX 677
Qy 668 SLERMAPLVRQVAGDVAYRPMAPNFDSTAFIAQELILSGAQOQNGYTEPIILHRREE 727
Db 678 IMKMAAVVDTQAGDPAYLPMSDFDGSVAFQAVELVLKREQNGYTEPIVLRHRRLE 737
Qy 728 FKARAA 733
Db 738 LKAKQA 743

RESULT 5
Q937W7 PRELIMINARY; PRT; 723 AA.
ID AC Q937W7
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Malate synthase G.
GN MASG.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF39;
RA Garcia de los Santos A., Hynes M.F.;
RL "Malate synthase gene from Rhizobium leguminosarum.";
DR Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY059637; AAL17965.1;
DR InterPro; IPR001465; Malate_synthase.
DR InterPro; IPR006253; Malate_synthase.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRfams; TIGR01345; malate syn. G; 1.
SQ SEQUENCE 723 AA; 79677 MW; 4E879906CFD64444 CRC64;

Query Match 39.18; Score 2367.5; DB 2; Length 723;
Best Local Similarity 64.18; Pred. No. 9e-129;
Matches 464; Conservative 82; Mismatches 173; Indels 5; Gaps 3;

```

```

Qy 11 RVSVGNLRIAEVLVDFYNNALPGTDIDPDSFWAGVDKVVADLTPOCALLNARDELQAO 70
Db 3 RVDKXGAIETVLHDFLVEEVLPGLAVDADKFPADPSAIVHDLAPKNCALLAKRDELQVK 62
Qy 71 IDKWHRRVIEPIDMDAYRFLTEIGYLLPEPDDFTITTSVDABITTTAGPOLVVPVLN 130
Db 63 IDDWRRHG-APADMDEYQSFLREIGYLLPEGSDFOVSTQVDPPEIASIAGPOLVVPVMN 121
Qy 131 APFALNANRWGSLYDALYGTDAIPETDGAERGRVNPGRGKVIAMARNFLDESA 190
Db 122 ARYALNANRWGSLYDALYGTDAIPESDGAERGRVNPGRGKVIAMARNFLDESA 181
Qy 191 SGSGDAGTGTVDQGVQVLPALPKSGTGLANPGFAGYTGABESPTSVLLINHLGHIE 249
Db 182 DCRMKDVGSAFVKGALVRSIDGQEAULTDGHFAGYRGDAAAAPHILLKNGIHIEV 241
Qy 250 IDPESQVGTTRDRAGVKDVILESATTTIMDFEDSVAADADKVLGYRNWLGKNGDLAAA 309
Db 242 IDAATTICKADSAHISDWLESATTTIMDCEDSIAADKVKVVRNWLGLKNGDLQEE 301
Qy 310 VDKGTAFRLVLRNDRNTYAPGGQFTLPGRSLMPFVRNVGHMTNDIAIVTDGSEVFE 369
Db 302 VAKGTSFIRTLNPDLOYAGPDGAFAFVHRSLMLVRNVGHMTNPAILDGRDQDVE 361
Qy 370 MDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMGHGAFAVFTCELFPSRVEDVL 429
Db 362 MDAATIGLIALYDICPS---GRRKNSRTGSIYIVKPKMGHGAFAVFAVEIFSRVEDALGL 418
Qy 430 PONTMKIGIMDEERTTIVNLKACIKAAADRVFINTGFLDRTGDEIHTSMEAGPMVRKGT 489
Db 419 PRNTMKGIMDEERTTIVNLKESIRAAKDRVVFINTGFLDRTGDEIHTSMEAGPMIRKGD 478
Qy 490 MKSOFWILAYEDHNVDAAGLAFSGRAQVKGKMTNTELMADVETKIAQPRAGASTAW 549
Db 479 MRQAAMIAAYENWNVDIGLECGLSGHAQIGKMWMPDLMAAMLEQKIAHPKAGANTAW 538
Qy 550 PSPTAATLHLYHGVDAVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPOBIREVDN 609
Db 539 PSPTAATLHLYHGVDAVAAVQOGLKRRAKLSLISVPAVPRNPTPEIQEELDNNA 598
Qy 610 QSILGYVVRWVQGVGCKVDPIDHVALMEDRATLRISSQLLANLWLRHGVITSADVRA 669
Db 599 QGILGYVVRWVQGVGCKVDPIDNIGLMDRATLRISSAQMANKLWLRHGVVTEAQIIE 658
Qy 670 ERMAPLVRQVAGDVAYRPMAPNFDSTAFIAQELILSGAQOQNGYTEPIILHRRRREFK 729
Db 659 RMAAVVDRQNASDPAYRPMAGNFDSDIAFQAALDLVLKREQNGYTEPIVLRHRRLE 718
Qy 730 ARAA 733
Db 719 AKQA 722

RESULT 6
Q92TA4 PRELIMINARY; PRT; 723 AA.
ID AC Q92TA4
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Probable malate synthase G protein (EC 4.1.3.2).
GN GLCB OR R00062 OR SMC02581.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,

```


RESULT 7
Q8YIR3

```

Db 547 SPTAATLHATHYHKIDVAAVQEKLSRPRAKLDDILSVFVAVRPNTWPTDDIQHEIDNNAQ 606
QY 611 SILGYVVRVWDQGVGSKVPDIHDVALMEDRATLRISOLLANLWLRHGVITSADVRASLE 670
Db 607 GILGYVVRVWDQGVGSKVPDINNVLGMDRATLRISAQHIANLWLVHGVVSEAQVWETMK 666
QY 671 RMAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGAQONGYTEPILHRRREFK 729
Db 667 RMAAIVDKQEGDPLRPMADFKSIAPQACDLVFKGREQNGYTEPVLHRRLELK 725

RESULT 8
Q8FZ50 PRELIMINARY; PRT; 728 AA.
AC Q8FZ50;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Malate synthase G.
GN GLCB OR BR1648.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=122711122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Unayam L., Brinkac L.M., Bean M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.B., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR EMBL; AE014458; AAN30550.1; -.
DR TIGR; BR1648; -.
KW Complete proteome.
SQ SEQUENCE 728 AA; 79966 MW; F95669D002A14EDE CRC64;

Query Match 38.2%; Score 2313; DB 16; Length 728;
Best Local Similarity 61.6%; Pred. No. 1.3e-125;
Matches 443; Conservative 97; Mismatches 175; Indels 4; Gaps 2;

QY 12 VSVGNLRIARVLYDFVNEALPGTDIDPDSFWAGVKVADLTPOQALLNARDELQAI 71
Db 10 VEIEGLAVAPELVEFLAKEAPGTGVEPEKFWGFAAIIERDLAPKRALAKEDELQARI 69
QY 72 DKHRRRVIEPIDMAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVWPVNA 131
Db 70 DAWYKENDKYSADYQQLKQILGILLPEGGAFTSVSTNVDPETHIAGPQLVWPVNA 129
QY 132 RFALNAANARWGLSYDALYGTVDVTPETDGAKEGPTYNKVRGDKVIAVARKFLDSDVPLS 191
Db 130 RYALNAANARWGLSYDALYGTDAISEADGAEGKGYNPKRGEKVIWAKNFLDESAPLST 189
QY 192 GSGGDATGFTVQDQGLVVALPKDS-TGLANPGOPAGVTGAESPTSVLLINHLGHTEILI 250
Db 190 GKWADVAGLVNDKGLSIRLTDGSAITLKDESQFKYNGDAAFTNVLLAKHMHVDIVI 249
QY 251 DPESQVGTDRAGVKDVIKESAITTMDPDSVAADAADKVLGYRNWLGLNKGDIAAAV 310
Db 250 NADHPIGKTDPAHIAADVLESASTIQDCEDSIAAADAEDKVAVYRNWLGLMNGKLEDTF 309
QY 311 DKDGTAFVLVLRNRYNTAPGGQFTLPGRSLMFPVNVGHLMNTDAIVDTDGSSEVEGIM 370
Db 310 EKNQKQMTNRUNGORTYTADPGSTLLTKHSLMLVRNVGHLMNTNPAILDAGNEVEGIM 369
QY 371 DALFTGLTIAHGLKASDVNGPLINSRTSGSIYIVKPKMHGPAEVAFTCELSRVEDVLGLP 430
Db 370 DAAFTSLIALHDI--GPNCRHMNSREGSVYIVKPKMHGPEEVAFAVEIIFTRTEMLGKM 426

```

```

QY 431 QNTMKIGIMDERRTVNLKACIKAAADRVVVFINTGFLDRTGDEIHTSMEAGPMVRKGTM 490
Db 427 PNTLKIGIMDERRTVNLKEAIRAAKDRVVVFINTGFLDRTGDEIHTSMEAGPMIRKGD 486
QY 491 KSQWILAYEDHNVNDAAGAFSGAQVQGVGMWMTMELMADMYETKTAQPRAGASTAWVP 550
Db 487 KOAWIGAYEQWNVDIIGLECCLSGHAQIGKGMWMPDMAANLEQKTAHPKAGANTAWVP 546
QY 551 SPTAATLHATHYHVDVAAVQOQGLAGKRRATIEOLLTIPLAKELAMAPDEIREEVNCCQ 610
Db 547 SPTAATLHATHYHKIDVAAVQEKLSRPRAKLDDILSVFVAVRPNTWPTDDIQHEIDNNAQ 606
QY 611 SILGYVVRVWDQGVGSKVPDIHDVALMEDRATLRISOLLANLWLRHGVITSADVRASLE 670
Db 607 GILGYVVRVWDQGVGSKVPDINNVLGMDRATLRISAQHIANLWLVHGVVSEAQVWETMK 666
QY 671 RMAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGAQONGYTEPILHRRREFK 729
Db 667 RMAAIVDKQEGDPLRPMADFKSIAPQACDLVFKGREQNGYTEPVLHRRLELK 725

RESULT 9
Q8FNB3 PRELIMINARY; PRT; 748 AA.
AC Q8FNB3;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Malate synthase (EC 4.1.3.2).
GN MAS2 OR CB2231.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005221; BAC19041.1; -.
KW Lyase; Complete proteome.
SQ SEQUENCE 748 AA; 83491 MW; F9550473BC4E9A09 CRC64;

Query Match 37.5%; Score 2273; DB 16; Length 748;
Best Local Similarity 61.5%; Pred. No. 2.9e-123;
Matches 444; Conservative 90; Mismatches 182; Indels 6; Gaps 4;

QY 9 TDRVSVGNLRIARVLYDFVNEALPGTDIDPDSFWAGVKVADLTPOQALLNARDELQ 68
Db 26 TERVTVGGMQVAKVLRDLFTESVLPVRGVDAERFWNGFGDIVRDMTPRNRELLARRDELQ 85
QY 69 AQIDKWHRRRVIEPIDMAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVWPV 128
Db 86 AQLODSYVRENEPKP-DPEKYEAFLEAFLGYLVDPEAPAEIRTONIDSEIATTAGPQLVVP 144
QY 129 LNARFALNAANARWGLSYDALYGTVDVTPETDGAKEGPTYNKVRGDKVIAVARKFLDSDV 188
Db 145 LNARFALNAANARWGLSYDALYGTNAIPDEDAERGAEPYFVRQKVIQNGRDELDVLP 204
QY 189 JSSGSGFGDATGFTVQDQGLVVALPKDSTGLANPGOPAGVTGAESPTSVLLINHLGHTEI 248
Db 205 LDGASHADVVEKYNITDGKLAHNVNDGIYRLKDRDAYLGTGTYFDPDPTSIQQNNGHIEL 264
QY 249 LTDPSQVGTDRAGVKDVIKESAITTMDPDSVAADAADKVLGYRNWLGLNKGDIAA 308
Db 265 QIDPHTPIKEDKTKGDKDIIIESAITTMDPDSVAADAEDKTLGYRNWFLNTNGELTE 324
QY 309 AYDKDGTAFVLVLRNRYNTAPGGQFTLPGRSLMFPVNVGHLMNTDAIVDTDGSSEVEG 368

```

```

Db 325 EVAKGDRFTFRKLNDRVFIKNGAELTLHGKSLLFVRNVGHMLMTNPAIL-VDGEEIYEG 383
QY 369 IMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKQHGPAEVAFTCELFSEVEDVLG 428
Db 384 IMDALIITTVCAIPGAPQKKK--NSRKGSIYIVKPKQHGPEEVAFTNELFAEVEDLLD 440
QY 429 LPQNTMKIGIMDEERTTNLAKACIKAAADRWVFINTGFLDRTGDEIHTSMEAGPMVRKG 488
Db 441 LPRHTLKGVGMDERTSVNLDAIMEVADRLAFINTGFLDRTGDEIHTSMEAGAVRKA 500
QY 489 TKMSOPWILAYEDHNVDAGLAGSGRAQVKGKGMWMTMELMADWVETKIAOPRAGASTAW 549
Db 501 DMOATAPWQAYEDNNVDAGIQRGUPGKACIQKGMWMTMELMGELEKKIGIQLEGANTAW 560
QY 549 VPSPTAATLHALHYHVDVAAVQOGL-AGKERATEQLLTTLPLAKELAWAPDEIREEDVN 607
Db 561 VPSPTGATLHATHYHRVDVFKVQDELRAAGRDSLSGLKLDVPVAPDTNWTDAEKRELDN 620
QY 608 NCQSTILGVVVRWVGCGSKVPDIHDVALMEDRATLRISQLLANLRLHGVITSADVRA 667
Db 621 NCQSTILGVVVRWVEGCGSKVPDIHDIDLMDRATLRISQILANLRLHGVITVEEQVIE 680
QY 666 SLERVAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAOPNGYTEPIHLRRRE 727
Db 681 SLERVAVVDQONAGDPNLYMAPNFTESVAFQARDLILKGTSPAGYITEPIHLRRRE 740
QY 728 FK 729
Db 741 FK 742

RESULT 10
Q9KB03
ID Q9KB03 PRELIMINARY; PRT; 727 AA.
AC Q9KB03
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Malate synthase.
GN BH2133.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirana C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331 (2000).
DR EMBL; AP001514; BAB05852.1; -.
DR HSSP; P37330; 1D8C.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
SQ Complete proteome.
KW SEQUENCE 727 AA; 81208 MW; CE18B45930E5AE5 CRC64;

Query Match
Best Local Similarity 37.3%; Score 2258; DB 16; Length 727;
Matches 441; Conservative 97; Mismatches 166; Indels 12; Gaps 4;

QY 1 MOHHHHHTDRVSGNLRVLYDFVNEALPGTDIDPDPSFWAGVDKVKVADLTPQNAL 60
Db 1 METYEH-----VGNLQVATKLITFIEQRALPGTGKIKDDFWSGVQQLITELMPENKML 53
QY 61 LNARDELQADIKHRRVRIEPIDMDAYRQFLTEIGYLLPEPDDFTTITTSVDAEITTA 120
Db 54 LAKREEIQADIMHQKNK-GPIDFSAYHSFLEIGYLEP:PEHVITITENVVDDEIAQA 112

```

```

QY 121 GPOLVVPVLNARPALNANRWGSLYDALYGTGVIPETDGAEGKPTYNKVRGDKVIAYAR 180
Db 113 GPOLVVPVNNARVAINANRWGSLYDALYGSNVISEEDGCEKGTGYNPKRGTKVIOFAK 172
QY 181 KFDDSDVPSGSGFGDATGFTVQDQGLVVALPD-KSTGLANPQGFAGYTGAAESPTSVLL 239
Db 173 DFLDHTFPLTSGSHHEALINAKMDQLVLTSLGSKWTRLKDETFQVYQSGQDPSVILL 232
QY 240 INHGLHIELIDPESQVITDRAGVKDVLIESAITTIMDFEDSVAADKVLGVNWL 299
Db 233 LHGLHVEIQIDARHPIGKSDRAKVDIVLESALITIMDCEDSVAADKEDKVVYQNL 292
QY 300 GLNKGDLAAAVDKGTAFLVLRNDRNYTAPGGQGTFLPGRSLMFRVNVGHMLTNDIYD 359
Db 293 GLMKGTLEATFFKEGKTKKLNEDRSYTAPNGETPSLGRSLMFRVNVGHMLTTPVIRT 352
QY 360 TDGSEVFEGIMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKQHGPAEVAFTCEL 419
Db 353 QSGEEVPEGILDGIVTSLIAKHDLQ--NGTFRNSKGSVIVKPKHGSBEVAFANRL 409
QY 420 FSRVEDVLGLPQNTMKIGIMDEERTTNLAKACIKAAADRWVFINTGFLDRTGDEIHTSM 479
Db 410 FNRIEDILGLERTHLKLGWDBERRISLNLKACIEKVKERVVFINTGFLDRTGDEIHTSM 469
QY 480 EAGPMVRKGTMSQOPWILAYEDHNVDAGLAGSGRAQVKGKGMWMTMELMADWVETKIAQ 539
Db 470 EAGPMIRKGDMSKSSWLSAYERSNVAAGLTCGFGRAQIKGMWAMPDLNMEWMEQKGTQ 529
QY 540 PRAGASTAWPSPATAATHALHYHVDVAAVQOGLAGKSRATIEQLLTPLAKELAWAPD 599
Db 530 LEAGANTAWPSPATAATHALHYHRRHVPAIQKTLADDQTCYRKKEILEPLAGSTDTWNE 589
QY 600 EIREEDVNNQSTILGVVVRWVGCGSKVPDIHDVALMEDRATLRISQLLANLRLHGV 659
Db 590 DIQAEELNNAQGLIGYVVRWVEGCGSKVPDIHNTALMEDRATLRISQHANLRLHGI 649
QY 660 ITSADVRASLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAOPNGYTEP 719
Db 650 VSXEQVIRTMERMAKVVDQONAGDPAYRPMADNLEQSVAFQALELVKGTQEPSGYTEP 709
QY 720 ILHRRRREFKARAAEK 735
Db 710 ILHRRLEFKQIAKE 725

RESULT 11
Q8FDN6
ID Q8FDN6 PRELIMINARY; PRT; 723 AA.
AC Q8FDN6
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Malate synthase G (EC 4.1.3.2).
GN GLCB OR C3705.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=06:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AE016766; AA82151.1; -.
KW Lyase; Complete proteome.
SQ SEQUENCE 723 AA; 80440 MW; AAF740B5FE038F6F CRC64;

```

Query Match 35.8%; Score 2168.5; DB 16; Length 723;
Best Local Similarity 58.4%; Pred. No. 3.1e-117;
Matches 422; Conservative 113; Mismatches 173; Indels 15; Gaps 5;

QY 12 VSVNLRIRARVLDFVNNALPGTDIDPDSFWAGVDKVVADLTTPQNALNARDLQAI 71
DB 5 ITQGLRIDANPKRFVDEEVLPGVELDAFAAHVHNDVDEIVHDLAPENRQLLADRDRIOAAL 64
QY 72 DKHRRVRIEPIIDMAYRQPLTEIGYLLPEPDDFTITTSQVDAEITTTAGPOLVVPVNA 131
DB 65 DEWHRNPGSKVKAAYKFLRELGLVLPQPDRTVETTGIDSEITTSQAGPOLVVPVNA 124
QY 132 RFLNANANRWGLYDALYGTVDVPTDGAEGKPTYNKVRGDKVAYARKPLDDSDVPLSS 191
DB 125 RYALNANANRWGLYDALYGSDDIIPQEGAVSG--YDPQGEQVIAWVRFIDESLPLEN 182
QY 192 GSGDATFTVQDQLVVALPD-KSTGLANPGQFAGYTGAAESPTSVLLINHLHIEILI 250
DB 183 GSYQDVVAFKVDKQLRIQLKNGKETTLRTPAQFVGYRGDTAAPTICILKXNGLHIEIQI 242
QY 251 DPSSOVGTTDRAGVKDVILESATTTIMDPEDSVAADKVLGYRNWGLNKGDLAAAV 310
DB 243 DANGRIKDSAHINDVIVAEALSTIIDCEDSVAAVDAEDKILLYRNLLGLMQGTLOEKN 302
QY 311 DKDGTAFRLVNRDRNTAPGGQFTLPGRSLMFVRNVGHMTNDAIVDTDGEVPEGIM 370
DB 303 EKNGRQIVRKLNDRDQVTAADGSEISLHGRSLFIRNVGELMTIPVIMWSEGEIPEGIL 362
QY 371 DALFTGLIAHGLKASDVNGPLNSRSTGYIVKPKMGPAEVAFTCELFSEVEDVLGLP 430
DB 363 DGVTGAIALYDLKVOX-----NSRTGSYIVKPKMGPAEVAFTCELFSEVEDVLGLP 416
QY 431 QNTWKIGIMDEERTTNLKAICAKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKGM 490
DB 417 PNLKXGIMDEERTSLNLRSCIAQARNVAFINTGFLDRTGDEMHSVVEAGPMLRKQM 476
QY 491 KSPFWILAYEDHNVDAAGLAFSGRAQVKGKMTWELMADVETKIAIPRAGASTAWVP 550
DB 477 KSTFWIKAYERNVLSGLFCGLRGKQIGKGMWMPDLMDMYSQKGDQLRAGANTAWVP 536
QY 551 SPTAATLHALHYHQVDVAAVVOOGLA-----GKRATIEOLLTIPLAKELAWAPDEIREV 605
DB 537 SPTAATLHALHYHQVNVQSVQANIAQTEFNAEPEPLDOLLITPVAENANWSVEHILQEL 596
QY 606 DNNQSTILGVYVRVQGVGCKVPDIHVALMEDRATLRISQLLANWLHGVITSADV 665
DB 597 DNNVQIGLYVVRVQGVGCKVPDIHVALMEDRATLRISQHLANWLHGVITSADV 656
QY 666 RASLERMAPVDONAGDVAYRPMANFDDSIAPLAAQELILSGAQQNGYTEPILHR-R 724
DB 657 QASLENKAVVDOONAGDPAYRPMAGNFANSCAFKASDLIFLGVKQNGYTEPILHAWR 716
725 RRE 727
717 LRE 719

RESULT 12

Q9CB77 PRELIMINARY; PRT; 421 AA.
ID AC Q9CB77
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Aspartokinase.
GN ASK OR ML2323.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.

RESULT 13

Q8KQ29 PRELIMINARY; PRT; 421 AA.
ID AC Q8KQ29
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Aspartate kinase alpha subunit (EC 2.7.2.4) (Aspartokinase).
GN ASKA.
OS Streptomyces clavuligerus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1901;
RN [1]
RP SEQUENCE FROM N.A.

Tunca S;
RA "Isolation and characterization of the aspartokinase and aspartate
RT semialdehyde dehydrogenase genes from cephamycin C-producer
RT Streptomyces clavuligerus";
RL Thesis (2002), University of Middle East Technical University, Ankara,
Turkey.
CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
CC ASPARTATE.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
DR EMBL; AV112728; AM54736.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate_kinase.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001057; Glu_kinase.
DR Pfam; PF00696; aakinas; 1.
DR Pfam; PF01842; ACT; 2.
DR PRINTS; PRO0474; GLUSKINASE.
DR TIGRFAMS; TIGR00657; asp_kinases; 1.
DR TIGRFAMS; TIGR00656; asp_kin_monofn; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 421 AA; 44399 MW; 08262D81045735C1 CRC64;

Query Match 21.0%; Score 1272.5; DB 2; Length 421;
Best Local Similarity 66.1%; Pred. No. 1.1e-65;
Matches 257; Conservative 51; Mismatches 60; Indels 21; Gaps 4;

QY 749 VQKYGSSVADAEIRIRVAERIVATKQGNVNVVSVSAMGDTTDDLLDLAQVCPAPPPR 808
DB 5 VQKYGSSLESADRIKVAERIVATKAGNEVVVSVCSAMGDTTDELLDLAQVNPVPPER 64

QY 809 ELDMLLTAGERISNALVMAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGRLQTA 868
DB 65 EMDMLLTAGERISNALVMAIAAQAQAWSFTGSGAGVITTTAVHGNARIIDVTPSRVTEA 124

QY 869 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFIFSADP 928
DB 125 LEEGVVALVAGFGVQADTKDITTLGRGSGDTTAVALAALNADVCETYSVDGVYSADP 184

QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYAREHNPVHVRSSYSDRPGTVVGS 988
DB 185 RIVPDARKLDSVPYEEMLELAASGSKILHRSVEYARYGVPIRVSSYSKPGTTVTGS 244

QY 989 IKDVPWEDPILTVGAHDSRSEAKVTIVGLPDIPGYAAKVFAVARRRRQHRHGAEE----- 1043
DB 245 IEEIPVEQALITVGAHDSRSEAKITVGVDPDHAGAAARIFRVA-----DAEIDIM 295

QY 1044 RLQG-----GRGDRHLLHLLPOTSGPPPKWTRSETRSASTQLLYDDHIGKVSLLIGAG 1098
DB 296 VLQVNTSSGRTD--ITFTLSKANGPKAVASLEKIEELGFSSVLYDDHVGKVSLLVGAG 353

QY 1099 MRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 354 MRSHPGVTATFCEALSKVGVNIELINTSE 382

RESULT 14
QYRQ25 PRELIMINARY; PRT; 421 AA.
AC QYRQ25;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 23, Last annotation update)
DE Aspartokinase subunit A (EC 2.7.2.4) (Aspartate kinase).
GN ASKA.
OS Amycolatopsis mediterranei (Nocardia mediterranei).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardineae; Pseudonocardiaaceae; Amycolatopsis.
OX NCBI_TaxID=33910;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=U32;

MEDLINE=99453302; PubMed=10521665;
RA Zhang W., Jiang W., Zhao G., Yang Y., Chiao J.;
RT "Sequence analysis and expression of the aspartokinase and aspartate
RT semialdehyde dehydrogenase operon from rifamycin SV-producing
RT amycolatopsis mediterranei";
RL Gene 237:413-419(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
CC ASPARTATE.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
DR EMBL; AF134837; AAD49567.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate_kinase.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001057; Glu_kinase.
DR Pfam; PF00696; aakinas; 1.
DR Pfam; PF01842; ACT; 2.
DR PRINTS; PRO0474; GLUSKINASE.
DR TIGRFAMS; TIGR00657; asp_kinases; 1.
DR TIGRFAMS; TIGR00656; asp_kin_monofn; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 421 AA; 44393 MW; 633D9C2D023145E9 CRC64;

Query Match 20.9%; Score 1267.5; DB 2; Length 421;
Best Local Similarity 65.3%; Pred. No. 2.2e-65;
Matches 254; Conservative 53; Mismatches 61; Indels 21; Gaps 4;

QY 749 VQKYGSSVADAEIRIRVAERIVATKQGNVNVVSVSAMGDTTDDLLDLAQVCPAPPPR 808
DB 5 VQKYGSSLESADRIKVAERIVATKAGNNDVVVSVCSAMGDTTDELLDLAQVNPAPPER 64

QY 809 ELDMLLTAGERISNALVMAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGRLQTA 868
DB 65 EMDMLLTAGERISNALVMAIAAQAQAWSFTGSGAGVITTVHGNARIIDVTPSRVTEA 124

QY 869 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFIFSADP 928
DB 125 LDQGYIALVAGFGVQADTKDITTLGRGSGDTTAVALAALNADVCETYSVDGVYTADP 184

QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYAREHNPVHVRSSYSDRPGTVVGS 988
DB 185 RIVPDARKLDTVTFEEMLELAASGSKILHRSVEYARYGVPIRVSSYSKPGTTVTGS 244

QY 989 IKDVPWEDPILTVGAHDSRSEAKVTIVGLPDIPGYAAKVFAVARRRRQHRHGAEE----- 1043
DB 245 IEEIPVEQALITVGAHDSRSEAKITVGVDPDHAGAAARIFRVA-----DAEIDIM 295

QY 1044 RLQG-----RGDRHLLHLLPOTSGPPPKWTRSETRSASTQLLYDDHIGKVSLLIGAG 1098
DB 296 VLQVNTSSGRTD--ITFTLSKANGAKAVKELEKVAEIGFESVLYDDHVGKVSLLVGAG 353

QY 1099 MRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 354 MRSHPGVTATFCEALAEAGVNIINTSE 382

RESULT 15
QYRQ25 PRELIMINARY; PRT; 421 AA.
AC QYRQ25;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Feedback-resistant aspartokinase LysC alpha subunit (EC 2.7.2.4)
GN ASKA.
OS Corynebacterium crenatum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=168810;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD945;

RA Liu Y., Ding J., Wang J.;
RT "Cloning and sequence analysis of aspartokinase genes from
RT *Cornebacterium crenatum*,"
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
CC -/- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
CC ASPARTATE.
CC -/- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.

DR EMBL; AF414084; AAL07807.1; -

DR InterPro; IPR001048; Aa_kinase.

DR InterPro; IPR002912; ACT.

DR InterPro; IPR001341; Aspartate_kinase.

DR InterPro; IPR005260; Asp_kin_monofn.

DR Pfam; PF00696; aakinese; 1.

DR Pfam; PF01842; ACT; 2.

DR TIGRFAMS; TIGR00657; asp_kinases; 1.

DR TIGRFAMS; TIGR00656; asp_kin_monofn; 1.

DR PROSITE; PS00324; ASPARTOKINASE; 1.

DR Kinase; Transferase.

KN SEQUENCE 421 AA; 44738 MW; 473A19409C0215E6 CRC64;

QY

Query Match 20.3%; Score 1232.5; DB 2; Length 421;

Best Local Similarity 64.3%; Pred. No. 2.3e-63;

Matches 252; Conservative 55; Mismatches 58; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRIRVAERIVATKQGNVWVVSAMGDTTDDLLDLAQVCPAPPER 808

Db 5 VQYGGSSLESARIRVAERIVATKAGNDVVVCSAMGDTTDELELAANVPVPPAR 64

QY 809 ELDMLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTGTHGNAKIIVTPGRLOTA 868

Db 65 EMDMLTAGERISNALVAMAEISLGAFAQSFQSGAGVLTTERHGNARIVDVTPGRVREA 124

QY 869 LEEGRVVLVAGFQGVSOQTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVGGIFSADP 928

Db 125 LDEGKICIVAGFQGVNKTREDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYADP 184

QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRLRCVEYARRHNI PVHVRSSYSDRPGTVVWGS 988

Db 185 RIVPNAQKLEKLEFEEMLELAARVGSKILVLRSEYARAFNVPLRVSSYSNDPGTILAGS 244

QY 989 IKDVPMEDPILTVADHRSKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038

Db 245 MEDIPVEEAVLTGVTDKSEAKVTVLGISDKPGEAAKVFRALADAEINIDMVLQNVSSYE 304

QY 1039 HGAERLQGRGROD-RHHLHL--POTSGPPPKNWTSETRSASTOLLYDDHIGKVSLLI 1095

Db 305 DGTDTITFTCPRSDGRAMEILKKPQVQG-----NW-----TNVLYDDGVGKVSLLV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGNIELISTSE 1127

Db 351 GAGWKSHPGVTABFMEALRDVNVNIELISTSE 382

Search completed: November 21, 2003, 16:08:15
Job time : 60.6698 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 59.515 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-54

Perfect score: 5072

Sequence: 1 MGHVHHVHVIDIGTPTSW.....RAWTEAVIGNRRQDSKESK 983

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5072	100.0	983	22 AAU01901	M. tuberculosis Tb
2	2700	53.2	802	19 AA081746	M. tuberculosis fu
3	2700	53.2	802	19 AA064379	Mycobacterium anti
4	2700	53.2	802	20 AA032063	Mycobacterium tube
5	2700	53.2	802	20 AA039224	M. tuberculosis fu
6	2700	53.2	802	20 AA0392176	M. tuberculosis fu
7	2700	53.2	802	20 AA039081	M. tuberculosis fus
8	2700	53.2	802	20 AA039033	M. tuberculosis fu
9	2700	53.2	802	23 AA074592	Antigenic fusion p

10	2375	46.8	460	19 AA081737	M. tuberculosis im
11	2375	46.8	460	19 AA064367	Mycobacterium tube
12	2375	46.8	460	20 AA039164	M. tuberculosis an
13	2375	46.8	460	20 AA039021	M. tuberculosis re
14	1839	36.3	652	20 AA039225	M. tuberculosis fu
15	1839	36.3	652	20 AA039082	M. tuberculosis fu
16	1831	36.1	374	19 AA081730	M. tuberculosis fu
17	1831	36.1	374	19 AA081731	M. tuberculosis 39
18	1831	36.1	374	19 AA064363	Mycobacterium anti
19	1831	36.1	374	19 AA064364	Mycobacterium 39 k
20	1831	36.1	374	20 AA032061	Mycobacterium tube
21	1831	36.1	374	20 AA039161	M. tuberculosis an
22	1831	36.1	374	20 AA039160	M. tuberculosis Tb
23	1831	36.1	374	20 AA039017	M. tuberculosis fu
24	1831	36.1	374	20 AA039018	M. tuberculosis 38
25	1831	36.1	374	22 AA011840	Mycobacterium tube
26	1831	36.1	374	23 AA029719	Mycobacterium sp.
27	1831	36.1	374	23 AA074590	Antigenic fusion p
28	1831	36.1	374	23 AA017593	Mycobacterium spec
29	1831	36.1	374	23 AA050733	Mycobacterium tube
30	1808	35.6	374	13 AA030090	38 kDa lipoprotein
31	791	15.6	166	18 AA032445	Mycobacterium tube
32	791	15.6	166	18 AA032377	Mycobacterium tube
33	791	15.6	166	19 AA081681	M. tuberculosis im
34	791	15.6	166	19 AA064319	Mycobacterium tube
35	791	15.6	166	20 AA039119	M. tuberculosis an
36	791	15.6	166	20 AA038982	M. tuberculosis re
37	485	9.6	100	19 AA081706	M. tuberculosis im
38	485	9.6	100	19 AA064339	Mycobacterium tube
39	485	9.6	100	20 AA039136	M. tuberculosis an
40	485	9.6	100	20 AA038993	M. tuberculosis re
41	485	9.6	100	20 AA035218	M. tuberculosis RV3
42	485	9.6	100	22 AA019845	Mycobacterium tube
43	485	9.6	100	23 AB005988	M. tuberculosis an
44	485	9.6	100	23 AB030969	Mycobacterium tube
45	481	9.5	100	20 AA030705	M. tuberculosis LH

ALIGNMENTS

RESULT 1

AAU01901
ID AAU01901 standard; Protein; 983 AA.

XX AAU01901;

AC AAU01901;

XX 29-AUG-2001 (first entry)

DT M. tuberculosis TbF15 fusion protein.

XX M. tuberculosis TbF15 fusion protein.

DE TbF15; antigen; vaccine; tuberculosis; AIDS; His tag; Ra3; 38kD;

XX 38-1; FL TbH4; acquired immunodeficiency disease.

KW Synthetic.

XX Mycobacterium tuberculosis.

OS Mycobacterium tuberculosis.

XX Key

FT Binding-site

FT Location/Qualifiers

FT 3..8

FT /label= Histidine tag

FT /note= "Nickel chelating region used for purifying

FT the fusion protein"

FT 9..74

FT /label= Ra3_region

FT 75..425

FT /label= 38kD_region

FT 426..524

FT /label= 38-1_region

FT 525..983

FT /label= FL_TbH4_region

XX WO200124820-A1.

PN

XX

CC protein is used in a method for inducing protective immunity against
CC tuberculosis (TB). This sequence can be formulated into vaccines
CC and/or pharmaceutical compositions for immunising against
CC M. tuberculosis infection or may be used for the diagnosis of TB.
XX
SQ Sequence 802 AA;
Query Match 53.2%; Score 2700; DB 19; Length 802;
Best Local Similarity 66.6%; Pred. No. 2.9e-158;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
QY 1 MGHHEHHVHDIIGTSPTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITRYI 60
Db 1 MGHHEHHVHDIIGTSPTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITRYI 60
QY 61 KLEVSFKMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGSTITLLYPLNLWG 120
Db 61 KLEVSFKMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGSTITLLYPLNLWG 119
QY 121 PAFHERYPNVTITTAQGTGSGAGIAQAAAGTGNIGASDAYLSEGDMAAHKGMLNIALAISA 180
Db 120 PAFHERYPNVTITTAQGTGSGAGIAQAAAGTGNIGASDAYLSEGDMAAHKGMLNIALAISA 179
QY 181 QQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239
QY 241 GDTFLFTQYLSKQDEGKPGFGFTTVDFFAVPGALGENGGMTVCATPCGVAYIG 300
Db 240 GDTFLFTQYLSKQDEGKPGFGFTTVDFFAVPGALGENGGMTVCATPCGVAYIG 299
QY 301 ISFLDQASQRLGEAQLGSSGNFLFLPDAQSIQAAAGFASKTPANQAISMIDGPAEDGY 360
Db 300 ISFLDQASQRLGEAQLGSSGNFLFLPDAQSIQAAAGFASKTPANQAISMIDGPAEDGY 359
QY 361 PIINYEYAVNRRQKDAATQTLQAFHLWATDGNKASFLDQVHFPQPLPRAVVKLSALI 420
Db 360 PIINYEYAVNRRQKDAATQTLQAFHLWATDGNKASFLDQVHFPQPLPRAVVKLSALI 419
QY 421 ATISSAENKTDAAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWREGAAGTAAQAAVV 480
Db 420 ATISSAENKTDAAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWREGAAGTAAQAAVV 479
QY 481 RFQEAANKQKBLDISINIRQAGVQYGRABEEOQQAALSSOMGFTQSTQTVVDQOEILNR 540
Db 480 RFQEAANKQKBLDISINIRQAGVQYGRABEEOQQAALSSOMGFV--PTTAAASPPSTAAA 537
QY 541 ANEVEAPNADPPTDVTIPCELTAAKNAQQVLVSADNMREYLAAGAKERQRLATSLENA 600
Db 538 PPAPATPVAPPPAAANTPNQCPGPNAPPPADPNAPPPVIAPNADQPVV----- 589
QY 601 AKAYGEVDEEATALDNDGEGTVQAESAGAVGGDS-----SABLTDTTPRVATAGEPNF- 653
Db 590 -----IDNPVGGSFSPALPAGWVESDAHPDYGSAIIS-----KTTGDPFPF 630
QY 654 -----MDLKEARLETQDQASLAHFDGNTFNLTLDQGVXFRFGPD 697
Db 631 GQPPVANDTRIVLRDLQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
QY 698 NWEQDAATACEASLDQQRQWILLMAKLSAAMAKAQYVVAQLFVWARREHPTVEDIVGLER 757
Db 676 PYPGTRINQETVSLD-----ANGVSGSASVYEVKFSKPNQKQWTVGIGSPA 724
QY 758 LYAENPSARDQLPVAEYQORSEKVLTEYNN-----KAALEVPNP-PKPPPAIKIDP 809
Db 725 ANAPDAGPPQRFVYV-----LGTANNFVDKGAALAESIRFLVAPPEA----P 770
QY 810 PPPPOEQGLIPGFLMPPSDGSGVTPGTGMPAAPVPTGSPGGGLPA 856
Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTFTFORTLPA 802

RESULT 3

AAW64379
ID AAW64379 standard; Protein; 802 AA.
XX
AC AAW64379;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium antigen TbF2 protein fusion.
XX
KW Tuberculosis; infection; diagnosis; 38 kDa antigen; TBra3; DPEP;
KW TB38-1; TbF-2.
XX
OS Mycobacterium tuberculosis.
OS Synthetic.
XX
PN WO9816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-251292/22.
DR N-PSDB; AAV55801.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Example 7; Page 223-226; 250pp; English.
XX
CC This polypeptide comprises a fusion protein, designated TbF-2,
CC composed of Mycobacterium tuberculosis antigens TBra3 (see AAW64295),
CC 38 kDa antigen (see AAW64364), Tb38-1 (see AAW64321) and DPEP (see
CC AAW64322). It was produced by PCR amplification (see AAV44450-57) of
CC the appropriate antigen DNA sequences, cloning into an expression
CC vector, and expression in E. coli. TbF-2 can be used for
CC serodiagnosis of tuberculosis. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising antigenic or
CC immunogenic portions of M. tuberculosis antigens, or fusion proteins,
CC DNA sequences encoding such polypeptides, recombinant expression
CC vectors and host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient.
SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 19; Length 802;
Best Local Similarity 66.6%; Pred. No. 2.9e-158;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
QY 1 MGHHEHHVHDIIGTSPTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITRYI 60
Db 1 MGHHEHHVHDIIGTSPTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITRYI 60
QY 61 KLEVSFKMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGSTITLLYPLNLWG 120
Db 61 KLEVSFKMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGSTITLLYPLNLWG 119
QY 121 PAFHERYPNVTITTAQGTGSGAGIAQAAAGTGNIGASDAYLSEGDMAAHKGMLNIALAISA 180
Db 120 PAFHERYPNVTITTAQGTGSGAGIAQAAAGTGNIGASDAYLSEGDMAAHKGMLNIALAISA 179
QY 181 QQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239

QY 654 -----MDLKEARKLETGDOGASLAHFADGWNFTNLTQGDVRRGFD 697
 Db 631 GPPPVANDTRIVLGRLDQKLYASAEATDSKAAA-----RLGSDMGEP--YM 675
 QY 698 NWEGDAATACASLDQOQOWILHMAKLSAAMAKQAQYVAQLHVWARRHEHTYEDIVGLER 757
 Db 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSKPKNGQIWTGVIGSPA 724
 QY 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
 Db 725 ANAPDAGPPQWFFVW-----LGTANNPVDKGAALAEIRPLVAPPPA----P 770
 QY 810 PPPQEQGLIFGLMPPSDGSGVTGTCMPAAPMVPTGSPGGGLPA 856
 Db 771 APAPAEPA-----PAPAPAGEVAP-----TFTPTPQRTLPA 802

RESULT 5
 ID AAY39224
 XX AAY39224 standard; Protein; 802 AA.
 AC AAY39224;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein Tbf-6 amino acid sequence.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Synthetic.
 OS Mycobacterium tuberculosis.
 XX
 XX WO9942076-A2.
 XX
 XX 26-AUG-1999.
 XX
 XX 17-FEB-1999; 98WO-US03268.
 XX
 XX 05-MAY-1998; 98US-0072967.
 XX 18-FEB-1998; 98US-0025197.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 XX WPI: 1999-527409/44.
 DR N-PSDB; AA219457.
 XX
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Claim 37; Page 271-273; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.
 XX
 SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 20; Length 802;
 Best Local Similarity 66.6%; Pred. No. 2.9e-158;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

QY 1 MGHVHHVVDIIGTSTWEQAAAEAVQARSDVDIRVARVIEODMAVDSAGKITRYI 60
 Db 1 MGHVHHVVDIIGTSTWEQAAAEAVQARSDVDIRVARVIEODMAVDSAGKITRYI 60
 QY 61 KLEVSFKMRPAQPCRCGSKPPSPETGAGAGTATTASSPVTTAETGTLTYPLENMG 120
 Db 61 KLEVSFKMRPAQPR-GSKPPSGSPETGAGAGTATTASSPVTTAETGTLTYPLENMG 119
 QY 121 PAFHERYPNVTTAQTGSGAGTAAAGTAVNGASDAYLSEGDMAAHKGLNIALAISA 180
 Db 120 PAFHERYPNVTTAQTGSGAGTAAAGTAVNGASDAYLSEGDMAAHKGLNIALAISA 179
 QY 181 QQVNYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS 240
 Db 180 QQVNYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS 239
 QY 241 GDTFLTQYLSKQDPGKSKSPGFGTTVPFPAVPGALGENGNGMTGCAETPGCVAYIG 300
 Db 240 GDTFLTQYLSKQDPGKSKSPGFGTTVPFPAVPGALGENGNGMTGCAETPGCVAYIG 299
 QY 301 ISFLDOASQRLGEAQLGNSSGNFLPDQAQSIQAAAAGFASKTPANCAISMIDGPADGY 360
 Db 300 ISFLDOASQRLGEAQLGNSSGNFLPDQAQSIQAAAAGFASKTPANCAISMIDGPADGY 359
 QY 361 PIINYEYAIVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKLSDALI 420
 Db 360 PIINYEYAIVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKLSDALI 419
 QY 421 ATSSAEMKTDAAATLAGEACNFERISGDLTKTDQVESTAGSLQOGWRGAAGTAAQAAVV 480
 Db 420 ATSSAEMKTDAAATLAGEACNFERISGDLTKTDQVESTAGSLQOGWRGAAGTAAQAAVV 479
 QY 481 RFOEAAKQKQELDEISTNIRQAGVQVSRADDEQQAALSSQMGFTQSTVTVDDQELNR 540
 Db 480 RFOEAAKQKQELDEISTNIRQAGVQVSRADDEQQAALSSQMGFV--PTTAAASPSTAAA 537
 QY 541 ANEVEAPMADPTDVTITPCELTAAKNAQQLVLSADNREYLAAGAKERQRLATSLRNA 600
 Db 538 PPAPATPVAPPPPPAAANTPNAPQDPNAPPPADPNAPPPPPVAPNAPQVPR-----589
 QY 601 AKAYGEVDEEAATALDNDGEGTVQAESAGAVGDS-----SALETDTPRVATAGBNF- 653
 Db 590 -----INDPVGFSFALPAGVSDAAHFDYGSALLS-----KITGDPFPP 630
 QY 654 -----MDLKEARKLETGDOGASLAHFADGWNFTNLTQGDVRRGFD 697
 Db 631 GPPPVANDTRIVLGRLDQKLYASAEATDSKAAA-----RLGSDMGEP--YM 675
 QY 698 NWEGDAATACASLDQOQOWILHMAKLSAAMAKQAQYVAQLHVWARRHEHTYEDIVGLER 757
 Db 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSKPKNGQIWTGVIGSPA 724
 QY 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
 Db 725 ANAPDAGPPQWFFVW-----LGTANNPVDKGAALAEIRPLVAPPPA----P 770
 QY 810 PPPQEQGLIFGLMPPSDGSGVTGTCMPAAPMVPTGSPGGGLPA 856
 Db 771 APAPAEPA-----PAPAPAGEVAP-----TFTPTPQRTLPA 802

RESULT 6
 AAY39176
 ID AAY39176 standard; Protein; 802 AA.
 XX
 XX AAY39176;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein Tbf-2 amino acid sequence.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;

KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.

OS Synthetic.
 OS Mycobacterium tuberculosis.

XX WO9942076-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03268.

XX PR 05-MAY-1998; 98US-0072967.

XX PR 18-FEB-1998; 98US-0025197.

XX PA (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;

XX WPI; 1999-527409/44.
 XX N-PSDB; AAZ19368.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions

XX Disclosure; Page 205-208; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T. B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.

XX SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 20; Length 802;
 Best Local Similarity 66.6%; Pred. No. 2.9e-158;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

QY 1 MGHHHHHVDDIIGTSTWEQAAABAVQARDSDVDIRVARVIEDQMAVDSAGKITVRI 60
 DB 1 MGHHHHHVDDIIGTSTWEQAAABAVQARDSDVDIRVARVIEDQMAVDSAGKITVRI 60
 QY 61 KLEVSFMRPAQRCQKPPSGSPETGAGATVATTASSPVTLAETGTLPLPLNFWG 120
 DB 61 KLEVSFMRPAQRCQKPPSGSPETGAGATVATTASSPVTLAETGTLPLPLNFWG 119
 QY 121 PAFHERYPNVITTAQTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNLALISA 180
 DB 120 PAFHERYPNVITTAQTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNLALISA 179
 QY 181 QQVYNLPGVSEHLKNGKVLAAVGGTITWDDPOIALNPGVNLPGTAVVPLHRSOGS 240
 DB 180 QQVYNLPGVSEHLKNGKVLAAVGGTITWDDPOIALNPGVNLPGTAVVPLHRSOGS 239
 QY 241 GDTFLEQYLSKQDPGKSGSPGFGTITVDFPAVPGALGNGGVMVTCAGETPGCVAYIG 300
 DB 240 GDTFLEQYLSKQDPGKSGSPGFGTITVDFPAVPGALGNGGVMVTCAGETPGCVAYIG 299
 QY 301 ISFLDAQSQRGLGEALGNSGNFLPLDQAQSTQAARAGFASKTPANQALSMIDGFPDGY 360
 DB 300 ISFLDAQSQRGLGEALGNSGNFLPLDQAQSTQAARAGFASKTPANQALSMIDGFPDGY 359
 QY 361 PIINYEYAVINNRKQDAATQTLQAEHLWAIIDGNKASFLDQVHFQPLPPAVVKLSDALI 420
 DB 360 PIINYEYAVINNRKQDAATQTLQAEHLWAIIDGNKASFLDQVHFQPLPPAVVKLSDALI 419

QY 421 ATISSAEMKTDAAATLAQAGNFERISGDLTKTQIDOVESTAGSLQGWGRGAAGTAAQAAVV 480
 DB 420 ATISSAEMKTDAAATLAQAGNFERISGDLTKTQIDOVESTAGSLQGWGRGAAGTAAQAAVV 479
 QY 481 RFQBAANKQKQELDEISTNIRQAGVQYSRADBEQQQALSSQMFTOSQTVTVQOEILNR 540
 DB 480 RFQBAANKQKQELDEISTNIRQAGVQYSRADBEQQQALSSQMFTOSQTVTVQOEILNR 537
 QY 541 ANEVEAPWADPPDTPVITPCELTAANKAAQOLVLSADNWEYLAAGAKERQRLATSLRNA 600
 DB 538 PPAPATPVAPPPFAAANTPNAQCDPNAPPPADPNAPPPVPIAPNAPQVPR----- 589
 QY 601 AKAYGEVDEEAATALDNDGEGTVQAESAGAVGSDS-----SAELTDTPRVATAGFNF- 653
 DB 590 -----IDNPGVGSFALPAGWVESDAAHFDYGSALLS-----KITGDPPPP 630
 QY 654 -----MDLKEAARKLETGQGLASLAHFADGWNFTNLTQGVKRRGFD 697
 DB 631 GQPPPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
 QY 698 NWEGDAATACEASLDQQRQWILHMAKLSAAMAKQAQVVAQLHVWARREHPTIEDIVGLER 757
 DB 676 PYPOTRINQETVSLD-----ANGVSGSASYEVKFSKPNQIWTGVIGSPA 724
 QY 758 LYAENPARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
 DB 725 ANAPDAGPPQRFVYVW-----LGTANNPVDKGAALAEISIRPLVAPPPA----P 770
 QY 810 PPPQEGLIPGLPMPSPDGSVTPGTGMPAAMPVPTGSPGGGLPA 856
 DB 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 7

AY39081

ID AAY39081 standard; Protein; 802 AA.

XX AAY39081;

XX 05-NOV-1999 (first entry)

XX M tuberculosis fusion protein Tbf-6.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.

XX Mycobacterium tuberculosis.

OS WO9942118-A2.

PN 26-AUG-1999.

PD 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

PA Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;

XX WPI; 1999-527416/44.

DR N-PSDB; AAZ19245.

XX New polypeptide comprising antigenic portions of M. tuberculosis

PS Example 10; Page 316-318; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a

CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.

XX
SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 20; Length 802;
Best Local Similarity 66.6%; Pred. No. 2.9e-158;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

Qy 1 MGHHEHHHVIDIIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYI 60
Db |||||
Qy 1 MGHHEHHHVIDIIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYI 60
Db |||||
Qy 61 KLEVSFMRPAQRCGSKPPSPETGAGACTVATTTPASSPVTLAETGTLTYLPLENLWG 120
Db |||||
Qy 61 KLEVSFMRPAQRCGSKPPSPETGAGACTVATTTPASSPVTLAETGTLTYLPLENLWG 119
Db |||||
Qy 121 PAFHERYPNVTITAGTSGGAGIAQAAGTGNICASDAYLSEGDMAHKGMLNIALAISA 180
Db |||||
Qy 120 PAFHERYPNVTITAGTSGGAGIAQAAGTGNICASDAYLSEGDMAHKGMLNIALAISA 179
Db |||||
Qy 181 QQVYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db |||||
Qy 180 QQVYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239
Db |||||
Qy 241 GDTFLFTQYLSKQDPEGKSGPGFTTVDFFPAVFGALGENGGWVTGCAETPCCVAYIG 300
Db |||||
Qy 240 GDTFLFTQYLSKQDPEGKSGPGFTTVDFFPAVFGALGENGGWVTGCAETPCCVAYIG 299
Db |||||
Qy 301 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPAPDGY 360
Db |||||
Qy 300 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPAPDGY 359
Db |||||
Qy 361 PIINYEIVANNRQKDAATQLOAFHWAITDGNKASFQDFQVHPQLPPLPAVKLSDALI 420
Db |||||
Qy 360 PIINYEIVANNRQKDAATQLOAFHWAITDGNKASFQDFQVHPQLPPLPAVKLSDALI 419
Db |||||
Qy 421 ATISSAEMKTDAAATLAQAGNFERISGLTKQIDOVESTAGSLOGWRGAAGTAAQAAV 480
Db |||||
Qy 420 ATISSAEMKTDAAATLAQAGNFERISGLTKQIDOVESTAGSLOGWRGAAGTAAQAAV 479
Db |||||
Qy 481 RFQEAANKQKQELDEISTNIRQAGVQYGRADDEEQQALSSQWGFQTSQTVTVDDQELNR 540
Db |||||
Qy 480 RFQEAANKQKQELDEISTNIRQAGVQYGRADDEEQQALSSQWGFV--PTTAASPPSTAAA 537
Db |||||
Qy 541 ANEVETAPVADPTDPTTPELTAAKNAQQVLSADNREYLAAGAKERORLATSUNA 600
Db |||||
Qy 538 PPAPATPVAPPPAAAANTPNAQPGDPNAAAPPADPNAPPVPIAPNAPQVPR----- 589
Db |||||
Qy 601 AKAYCEVDEEAATALDNDGEGTVQAESAGVGDGDS-----SAELTDTFVATAGEBNF- 653
Db |||||
Qy 590 -----IDNPVGGFSFALPAGWVESDAAHFDYGSALLS-----KTTGDBPFP 630
Db |||||
Qy 654 -----MDLKEARKLETGDDQASLAHFAQGNWTFNLTLQGDVVKRFRGFD 697
Db |||||
Qy 631 GQPPVPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGCF--YM 675
Db |||||
Qy 698 NWEGDAATACERASLDQQRWILHMAKLSAAMAKQAYVLAQLHVHARREHPYEDIVGLER 757
Db |||||
Qy 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDFSKPNQGTWTVGIGSPA 724
Db |||||
Qy 758 LYAENPSARDQILPVYAYQQRSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
Db |||||
Qy 725 ANAPDAGPPQRFVVM-----LGTANNFVDKGAAKALAESIRPLVAPPPA----P 770
Db |||||
Qy 810 PPPPQEQGLIPGFLNPPSDGSGVTTGTMPAAPVPPPTGSGPGGLUPA 856
Db |||||
Qy 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTTQRTLEPA 802
Db |||||

RESULT 8

AAV39033
ID AAV39033 standard; Protein; 802 AA.
XX AC AAV39033;
XX DT 05-NOV-1999 (first entry)
XX DE M. tuberculosis fusion protein TbF-2.
XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;
XX KW vaccine; immunity.
XX OS Mycobacterium tuberculosis.
XX PN W09942118-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WC-US03265.
XX PR 05-MAY-1998; 98US-0072596.
XX PR 18-FEB-1998; 98US-0024753.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Reto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;
XX WP1; 1999-527416/44.
XX DR N-PSDB; AA219156.
XX PT New polypeptide comprising antigenic portions of M. tuberculosis
XX PS Example 10; Page 251-253; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.

XX
SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 20; Length 802;
Best Local Similarity 66.6%; Pred. No. 2.9e-158;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

Qy 1 MGHHEHHHVIDIIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYI 60
Db |||||
Qy 61 KLEVSFMRPAQRCGSKPPSPETGAGACTVATTTPASSPVTLAETGTLTYLPLENLWG 120
Db |||||
Qy 61 KLEVSFMRPAQRCGSKPPSPETGAGACTVATTTPASSPVTLAETGTLTYLPLENLWG 119
Db |||||
Qy 121 PAFHERYPNVTITAGTSGGAGIAQAAGTGNICASDAYLSEGDMAHKGMLNIALAISA 180
Db |||||
Qy 120 PAFHERYPNVTITAGTSGGAGIAQAAGTGNICASDAYLSEGDMAHKGMLNIALAISA 179
Db |||||
Qy 181 QQVYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db |||||
Qy 180 QQVYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239
Db |||||
Qy 241 GDTFLFTQYLSKQDPEGKSGPGFTTVDFFPAVFGALGENGGWVTGCAETPCCVAYIG 300
Db |||||
Qy 240 GDTFLFTQYLSKQDPEGKSGPGFTTVDFFPAVFGALGENGGWVTGCAETPCCVAYIG 299
Db |||||
Qy 301 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPAPDGY 360
Db |||||
Qy 300 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPAPDGY 359
Db |||||

QY 361 PIINYEYAI VNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKSALDI 420
 Db 360 PIINYEYAI VNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKSALDI 419
 QY 421 ATISSAEMKTDAAATLAQEAAGNFERISGDLTKTQIDQVESTAGSLQGWGAAAGTAAQAAV 480
 Db 420 ATISSAEMKTDAAATLAQEAAGNFERISGDLTKTQIDQVESTAGSLQGWGAAAGTAAQAAV 479
 QY 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADEBQQQALSSQMGFTQSVTVDDQBEILNR 540
 Db 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADEBQQQALSSQMGFTV--PTTAAASPPSTAAA 537
 QY 541 ANEVEAPMADPPTDVTTPCELTAANKAAQQLVLSADNMREYLAAGAKERQRLATSLRNA 600
 Db 538 PPAPATFVAPPPAAANTFNAQCDPNAPPPADPNAPPPVIAFNAPOFVR-----589
 QY 601 AKAYGEVDEEAATALNDGEGTVQASAGAVGDS-----SAELTDTPRVATAGEPNF- 653
 Db 590 -----IDNPVGGSFALPAGWVESDAAHFDYGSALLS-----KTTGDPFPP 630
 QY 654 -----MDLKEAARKLETGQGNLAHFAFGWNTFNLTLQGVKFRGPD 697
 Db 631 GQPPPVANDTRIVLGRDQKLYASAEATDSKAAA-----RLGSDMGFE--YM 675
 QY 698 NWEGDAATACEASLDQOROWILHMAKLSAAMAKQAQYVAQLHVAWRREHPTVEDIVGLER 757
 Db 676 PYFGTRINQETVSLD-----ANGVSGSASYEVKFSDPKPNQIWTGVIGSPA 724
 QY 758 LYAENSARDQILPVVAEYQORSEKVLTEYNN-----KALEPVNP-PKPPPAIKIDP 809
 Db 725 ANAPDAGPQPRWFVW-----LGTANNPVDRGAALAESIRPLVAPPPA-----P 770
 QY 810 PPPPQOGLIPGLMPPSDGSGVTCTGMPAAPVMPPTGPGGGLPA 856
 Db 771 APAPAPPA-----PAPAPAGEVAP-----TPTTTPORTLPA 802

RESULT 9
 AAU74592
 ID AAU74592 standard; Protein; 802 AA.
 AC AAU74592;
 XX
 DT 08-MAY-2002 (first entry)
 DE Antigenic fusion protein Tbra3-38kd-Tb38-1-DPEP (TbF-2).
 DE Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; Tbra3-38kd-Tb38-1-DPEP; TbF-2.
 XX Chimeric - Mycobacterium tuberculosis.
 OS
 XX US2002009459-A1.
 XX
 XX 24-JAN-2002.
 XX
 XX 07-APR-1999; 99US-0287849.
 XX
 XX 13-MAR-1997; 97US-0818112.
 PR 01-OCT-1997; 97US-0942578.
 PR 18-FEB-1998; 98US-0025197.
 PR 07-APR-1998; 98US-0056556.
 PR 30-DEC-1998; 98US-0223040.
 XX
 XX (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMP/) CAMPOS-NETO A.
 XX
 PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 XX WPI; 2002-171134/22.
 DR

XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 PT diagnosing, treating or preventing M. tuberculosis infection,
 PT particularly as vaccine for treating or preventing tuberculosis -
 XX Claim 1; Fig 5G-J; 62pp; English.
 PS
 XX The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention.
 CC Note: The specification states that this polypeptide is encoded by the
 CC polynucleotide shown in ABK14132.
 XX
 SQ Sequence 802 AA;
 Query Match 53.2%; Score 2700; DB 23; Length 802;
 Best Local Similarity 66.6%; Pred. No. 2.9e-158;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
 QY 1 MGHHHHHVHDIIGTSPTEQAAAEAVQEARSDVDIRVARVIEODMAVDSAGKITRYI 60
 Db 1 MGHHHHHVHDIIGTSPTEQAAAEAVQEARSDVDIRVARVIEODMAVDSAGKITRYI 60
 QY 61 KLEVSFRRPAQRCSKPSGSPETGAGAGTATTTPASSPVTLAETGTLTYLFLNLWG 120
 Db 61 KLEVSFRRPAQR-CSKPSGSPETGAGAGTATTTPASSPVTLAETGTLTYLFLNLWG 119
 QY 121 PAFHERYPNTITTAQGTGSGAGIAQAAAGTNTGASDAYLSEGDMAHKLNMIALAISA 180
 Db 120 PAFHERYPNTITTAQGTGSGAGIAQAAAGTNTGASDAYLSEGDMAHKLNMIALAISA 179
 QY 181 QQVNNYLPVGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVFLHSDGS 240
 Db 180 QQVNNYLPVGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVFLHSDGS 239
 QY 241 GDTFLFTQYLSKQDPGEGKSPGFGTTVDPPAVPGALGENGNGMTGCAETGCVAYIG 300
 Db 240 GDTFLFTQYLSKQDPGEGKSPGFGTTVDPPAVPGALGENGNGMTGCAETGCVAYIG 299
 QY 301 ISFLDQASQSGLGEAQLGNSGNFLLPDAQSICAAAAGFASKTPANCAISMIDGPADGY 360
 Db 300 ISFLDQASQSGLGEAQLGNSGNFLLPDAQSICAAAAGFASKTPANCAISMIDGPADGY 359
 QY 361 PIINYEYAI VNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKSALDI 420
 Db 360 PIINYEYAI VNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKSALDI 419
 QY 421 ATISSAEMKTDAAATLAQEAAGNFERISGDLTKTQIDQVESTAGSLQGWGAAAGTAAQAAV 480
 Db 420 ATISSAEMKTDAAATLAQEAAGNFERISGDLTKTQIDQVESTAGSLQGWGAAAGTAAQAAV 479
 QY 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADEBQQQALSSQMGFTQSVTVDDQBEILNR 540
 Db 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADEBQQQALSSQMGFTV--PTTAAASPPSTAAA 537
 QY 541 ANEVEAPMADPPTDVTTPCELTAANKAAQQLVLSADNMREYLAAGAKERQRLATSLRNA 600
 Db 538 PPAPATFVAPPPAAANTFNAQCDPNAPPPADPNAPPPVIAFNAPOFVR-----589
 QY 601 AKAYGEVDEEAATALNDGEGTVQASAGAVGDS-----SAELTDTPRVATAGEPNF- 653
 Db 590 -----IDNPVGGSFALPAGWVESDAAHFDYGSALLS-----KTTGDPFPP 630
 QY 654 -----MDLKEAARKLETGQGNLAHFAFGWNTFNLTLQGVKFRGPD 697
 Db 631 GQPPPVANDTRIVLGRDQKLYASAEATDSKAAA-----RLGSDMGFE--YM 675

```

QY 698 NWEGDAATACASLDQORQWILHWAKLSAAVAKAQYVVAQLHVWARREHPTIEDIVGLER 757
D 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDFSKNGQIWTGVIQSPA 724
QY 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
D 725 ANAPDAGPPQRFVWV-----LGTANNPVVDKGAAKALAESIRPLVAPPPA-----P 770
QY 810 PPPPOEQGLIPGFLMPPSDGSGVTPGTGMPAAPVPPPTGSGPGGLIPA 856
D 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTILPA 802

RESULT 10
AAW81737
ID AAW81737 standard; Protein; 460 AA.
XX
AC AAW81737;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide TbH4-Xp1 #1.
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
FN WO9816646-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; 97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-261042/23.
DR N-PSDB; AAV64549.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
XX
PS Example 3c; Page 178-179; 230pp; English.
XX
CC This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
CC inducing protective immunity against tuberculosis (TB). This sequence can
CC be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis.
XX
SQ Sequence 460 AA;
Query Match 46.8%; Score 2375; DB 19; Length 460;
Best Local Similarity 100.0%; Pred. No. 1.6e-138;
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 525 TQSQTVVQOQELNRAVEAPMADPFDVITPCELTAANKAAQQLVLSADNNREYLA 584
D 2 TQSQTVVQOQELNRAVEAPMADPFDVITPCELTAANKAAQQLVLSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAKAYGEVDEEATALNDGEGTVQABSAGAVGDSAEITDTPR 644
D 62 AGAKERQRLATSLRNAKAYGEVDEEATALNDGEGTVQABSAGAVGDSAEITDTPR 121

```

```

QY 645 VATAGEPNMDLKEAARKLETGDOGASLAHPADGWNFTNLTOGDKVRFGRFDNWEGDAA 704
D 122 VATAGEPNMDLKEAARKLETGDOGASLAHPADGWNFTNLTOGDKVRFGRFDNWEGDAA 181
QY 705 TACEASLDQORQWILHWAKLSAAVAKAQYVVAQLHVWARREHPTIEDIVGLERLYAENPS 764
D 182 TACEASLDQORQWILHWAKLSAAVAKAQYVVAQLHVWARREHPTIEDIVGLERLYAENPS 241
QY 765 ARDQILPVYAEYQORSEKVLTEYNNKAALEFVNPPKPPPAIKIDPPPPPOEQGLIPGFLM 824
D 242 ARDQILPVYAEYQORSEKVLTEYNNKAALEFVNPPKPPPAIKIDPPPPPOEQGLIPGFLM 301
QY 825 PPSDGSQVTPGTGMPAAPVPPPTGSGPGGLIPADTAALQTSAGREAAALSGDVAVKAAASLG 884
D 302 PPSDGSQVTPGTGMPAAPVPPPTGSGPGGLIPADTAALQTSAGREAAALSGDVAVKAAASLG 361
QY 885 GGGGGVPSAPLGSAGIAESVAPAGAGDIAGLQGRAGGAGALGCGGGMPMGAHQOQ 944
D 362 GGGGGVPSAPLGSAGIAESVAPAGAGDIAGLQGRAGGAGALGCGGGMPMGAHQOQ 421
QY 945 GGAKSQSQEQDEALYTEDRANTEAVIGNRRRODSKESK 983
D 422 GGAKSQSQEQDEALYTEDRANTEAVIGNRRRODSKESK 460

RESULT 11
AAW64367
ID AAW64367 standard; Protein; 460 AA.
XX
AC AAW64367;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbH-4-Xp1.
XX
KW Tuberculosis; infection; diagnosis; antigen; TbH-4; Xp1.
XX
OS Mycobacterium tuberculosis strain H37Rv.
XX
OS Mycobacterium tuberculosis strain Erdman.
XX
FN WO9816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-251292/22.
DR N-PSDB; AAV44440.
XX
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Example 3; Page 189-191; 250pp; English.
XX
CC This polypeptide is the predicted amino acid sequence of
CC Mycobacterium tuberculosis antigen TbH4-Xp1. The sequence was
CC deduced from a composite sequence (see AAV44440) of overlapping DNA
CC clones TbH4 and Xp1. TbH4 (see also AAW64319) was isolated from a
CC M. tuberculosis strain H37Rv expression library and Xp1 from a M.
CC tuberculosis Erdman expression library. Recombinant Xp1 protein
CC was prepared. It stimulates cell proliferation and interferon-
CC gamma production in T cells isolated from M. tuberculosis-immune
CC donors. The invention relates to methods for diagnosing
CC tuberculosis. It provides polypeptides (see AAW64291-W64379)

```


KW vaccine; immunity.
XX Mycobacterium tuberculosis.
XX WO9942118-A2.
XX 26-AUG-1999.
XX 17-FEB-1999; 99WO-US02265.
XX 05-MAY-1998; 98US-0072596.
XX 18-FEB-1998; 98US-0024753.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527416/44.
XX N-PSDB; AA219138.
XX New polypeptide comprising antigenic portions of M. tuberculosis
PT Example 3; Page 219-220; 323pp; English.
XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX
SQ Sequence 460 AA;
Query Match 46.8%; Score 2375; DB 20; Length 460;
Best Local Similarity 100.0%; Pred. No. 1.6e-138; Indels 0; Gaps 0;
Matches 459; Conservative 0; Mismatches 0;
QY 525 TCSQTVTDVQDQILNRANEVEAPMDPTDVTTCETLTAANKAAQQLVLSADNNREYLA 584
Db 2 TCSQTVTDVQDQILNRANEVEAPMDPTDVTTCETLTAANKAAQQLVLSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAKAYGEVDEEATATLNDGEGTVQAESAGAGDSSAELTDTPR 644
Db 62 AGAKERQRLATSLRNAKAYGEVDEEATATLNDGEGTVQAESAGAGDSSAELTDTPR 121
QY 645 VATAGEPNFMDLKEARKLETGQASLAHFADGNTNLTQGDVKRFRGFDNWEGBAA 704
Db 122 VATAGEPNFMDLKEARKLETGQASLAHFADGNTNLTQGDVKRFRGFDNWEGBAA 181
QY 705 TACEASLDQORQWILHMAKLSAAMAKQAYVAQLHWARREHPTVEDIVGLERLYAENPS 764
Db 182 TACEASLDQORQWILHMAKLSAAMAKQAYVAQLHWARREHPTVEDIVGLERLYAENPS 241
QY 765 ARDQILPVYAEYQORSEKVLTEYNNKAALPEVNPBPKPAIKIDPPPPQOGLIPGLM 824
Db 242 ARDQILPVYAEYQORSEKVLTEYNNKAALPEVNPBPKPAIKIDPPPPQOGLIPGLM 301
QY 825 PPSDGSVTPGTGMPAAMPVPTSPGGLPADTAAQLTSAGREAAALSGLDVAVKAAASLG 884
Db 302 PPSDGSVTPGTGMPAAMPVPTSPGGLPADTAAQLTSAGREAAALSGLDVAVKAAASLG 361
QY 885 GGGGGVPSAPLGSAGAESVRPAGAGDIAGLGGGRAGGGAALGGGGMGPMGAHQOQ 944
Db 362 GGGGGVPSAPLGSAGAESVRPAGAGDIAGLGGGRAGGGAALGGGGMGPMGAHQOQ 421
QY 945 GGAKSQSGQDEEALYTEDRAWTEAVIGNRRQDSKESK 983
Db 422 GGAKSQSGQDEEALYTEDRAWTEAVIGNRRQDSKESK 460

RESULT 14

AA39225
ID AAY39225 standard; Protein; 652 AA.
XX AAY39225;
AC AAY39225;
XX 05-NOV-1999 (first entry)
XX M. tuberculosis fusion protein Tbf-8 amino acid sequence.
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX Synthetic.
OS Mycobacterium tuberculosis.
XX WO9942076-A2.
XX 26-AUG-1999.
XX 17-FEB-1999; 99WO-US03268.
XX 05-MAY-1998; 98US-0072967.
XX 18-FEB-1998; 98US-0025197.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527409/44.
XX N-PSDB; AA219460.
XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX Claim 37; Page 274-276; 299pp; English.
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 652 AA;
Query Match 36.3%; Score 1839; DB 20; Length 652;
Best Local Similarity 48.5%; Pred. No. 3.2e-105;
Matches 441; Conservative 26; Mismatches 132; Indels 310; Gaps 19;
QY 1 MGHHHHHVHDIIGTSTPSWEQAAAEAVQARSDVDIRVIEQDMAYDSAGKIYRI 60
Db 1 MGHHHHHH----- 8
QY 61 KLEVSFKMRPAQPRCGSKPPSPETGAGAGTVAATTPASSPVTLAETGTLTYPLFLNLG 120
Db 9 -----GSKPSPSPETGAGAGTVAATTPASSPVTLAETGTLTYPLFLNLG 53
QY 121 PAFHRYPNVTITTAQGTGSGAGIAAAAAGTVNIGASDAYLSEGDMAAHKLMMIALAISA 180
Db 54 PAFHRYPNVTITTAQGTGSGAGIAAAAAGTVNIGASDAYLSEGDMAAHKLMMIALAISA 113
QY 181 QQVNNYLPVSEHLKNGKVLAAVYOGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDDGS 240
Db 114 QQVNNYLPVSEHLKNGKVLAAVYOGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDDGS 173
QY 241 GDTFLFTQYLSKQDPGCKGKSPGFGTTVPFPAVPGALGENGNGMTVCATPCGVAYIG 300

Db 174 GDTFLFTQYLSKQDPGKSPGFTTVDPPAVFGALGENGGWVTGCAETPGCVAYIG 233
 Qy 301 ISFLDQASQRLGEAQLGNSGKFLPDQAQSIQAAAAGFASKTPANQAISMIDGPAPDGY 360
 Db 234 ISFLDQASQRLGEAQLGNSGKFLPDQAQSIQAAAAGFASKTPANQAISMIDGPAPDGY 293
 Qy 361 PIINYEYAI VNNRQKDAATQTLQAFLHWAITDGNKASFLDQVHFQPLPPAVVVKLSDALI 420
 Db 294 PIINYEYAI VNNRQKDAATQTLQAFLHWAITDGNKASFLDQVHFQPLPPAVVVKLSDALI 353
 Qy 421 ATISSAEMKTDATLAQEAENFERISGDLTKTDQVESTAGSLQGWGGAAGTAAQAAV 480
 Db 354 ATISSGG-----GSGGGSGGSGSVPTTAASPSTA-----AAPAPATPVA 396
 Qy 481 RFQEAANKQKQELDEISTNIRQAGVQVSRADDEQQQALSSQMGTQSTQVTVDDQQLNLR 540
 Db 397 PPPPAA-----ANTPNAQPG----- 411
 Qy 541 ANEVEAPMADPPTDVPITPCELTAAKNAQAQQLVLSADN-----MREYLAAGAKERQ----- 591
 Db 412 -----DENAAPPADPNAPPVPIAPNAPQPVRI--DNVGGFSFALPAGWVESDAAHFD 464
 Qy 592 -----RLATSLRNAAKAYGEVDEEAATALDND-- 618
 Db 465 YGSALLSKTTGDPFPFGQPPPVANDTRIVLGRDQKLYASAEA---TDSKAARLGSMDG 521
 Qy 619 -----GGTVOAESAGAVGSDSSAELTDTTPRVATAGEPNFMDLKEARKLETGD 667
 Db 522 EFYMPYPGTRINQETVSLDANGVSGSASYEV-----KFSDPKNGQIWTGV 569
 Qy 668 QGASLAHAFADGWTNLTLOGDVKFRGPDNWDGDAATACEASLDQORQWILEMAKLSAA 727
 Db 570 IGSPAANAPDA-----GPPQW--FVWILGTANNPVDK----- 600
 Qy 728 MAQQAQYVAQLHVHVARREHPTYEDIVGLERLYAENPSARDQILPVVAYEQORSEKVLTEY 787
 Db 601 -----GAAKALAE--SIRPLVAP----- 616
 Qy 788 NNKAALPEVNPVPPPAIKIDPPPPQBOGLIPGFLMPPSDSGSVTPGTCMPAAPVMPPT 847
 Db 617 -----PPAPAPA-PAEPAPAPAPAGEV-----AP-TPTT 643
 Qy 848 GSPGGGLPA 856
 Db 644 PTQRTILPA 652

RESULT 15

AAV39082
 ID AAV39082 standard; Protein; 652 AA.

AC AAV39082;

XX 05-NOV-1999 (first entry)

DE M tuberculosis fusion protein Tbf-8.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.

XX Mycobacterium tuberculosis.

XX WO942118-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99WO-US03265.

XX 05-MAY-1996; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX WPI; 1999-527416/44.
 DR N-PSDB; AAZ19248.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis
 PS Example 10; Page 320-321; 323pp; English.
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX
 SQ Sequence 652 AA;
 Query Match 36.3%; Score 1839; DB 20; Length 652;
 Best Local Similarity 48.5%; Pred. No. 3.2e-105;
 Matches 441; Conservative 26; Mismatches 132; Indels 310; Gaps 19;
 Qy 1 MGHHHHHHVDIIIGTSPTSWEQAAAAEAQVARDSDDIRVARVIEQDMAVDSAGKITRYI 60
 Db 1 MGHHHHH----- 8
 Qy 61 KLEVSFXMRPAQPCRGKPPSGSPETGAGAGTATVATTPASSPVTLAETGSLLYPLFNLWG 120
 Db 9 -----GSKPFGSPETGAGAGTATVATTPASSPVTLAETGSLLYPLFNLWG 53
 Qy 121 PAFHERVNVITTAQGTGSGAGIAQAAAAGTNNIGASDAYLSEGDMAHKLMMIALAISA 180
 Db 54 PAFHERVNVITTAQGTGSGAGIAQAAAAGTNNIGASDAYLSEGDMAHKLMMIALAISA 113
 Qy 181 QQVNNYLPVSEHLKNGKVLAAVYQGTIKTDDPQIAALNPGVNLPGTAVVPLHSDGS 240
 Db 114 QQVNNYLPVSEHLKNGKVLAAVYQGTIKTDDPQIAALNPGVNLPGTAVVPLHSDGS 173
 Qy 241 GDTFLFTQYLSKQDPGKSPGFTTVDPPAVFGALGENGGWVTGCAETPGCVAYIG 300
 Db 174 GDTFLFTQYLSKQDPGKSPGFTTVDPPAVFGALGENGGWVTGCAETPGCVAYIG 233
 Qy 301 ISFLDQASQRLGEAQLGNSGKFLPDQAQSIQAAAAGFASKTPANQAISMIDGPAPDGY 360
 Db 234 ISFLDQASQRLGEAQLGNSGKFLPDQAQSIQAAAAGFASKTPANQAISMIDGPAPDGY 293
 Qy 361 PIINYEYAI VNNRQKDAATQTLQAFLHWAITDGNKASFLDQVHFQPLPPAVVVKLSDALI 420
 Db 294 PIINYEYAI VNNRQKDAATQTLQAFLHWAITDGNKASFLDQVHFQPLPPAVVVKLSDALI 353
 Qy 421 ATISSAEMKTDATLAQEAENFERISGDLTKTDQVESTAGSLQGWGGAAGTAAQAAV 480
 Db 354 ATISSGG-----GSGGGSGGSGSVPTTAASPSTA-----AAPAPATPVA 396
 Qy 481 RFQEAANKQKQELDEISTNIRQAGVQVSRADDEQQQALSSQMGTQSTQVTVDDQQLNLR 540
 Db 397 PPPPAA-----ANTPNAQPG----- 411
 Qy 541 ANEVEAPMADPPTDVPITPCELTAAKNAQAQQLVLSADN-----MREYLAAGAKERQ----- 591
 Db 412 -----DENAAPPADPNAPPVPIAPNAPQPVRI--DNVGGFSFALPAGWVESDAAHFD 464
 Qy 592 -----RLATSLRNAAKAYGEVDEEAATALDND-- 618
 Db 465 YGSALLSKTTGDPFPFGQPPPVANDTRIVLGRDQKLYASAEA---TDSKAARLGSMDG 521
 Qy 619 -----GGTVOAESAGAVGSDSSAELTDTTPRVATAGEPNFMDLKEARKLETGD 667
 Db 522 EFYMPYPGTRINQETVSLDANGVSGSASYEV-----KFSDPKNGQIWTGV 569

QY	668	QASLAHFPADGNTFNLTLOQDVKFRGFDNWEGDAATACEASLDQOROWILHNAKLSAA	727
Db	570	IGSPAANAPDA-----GPPQKW--FVVMGTANNPVDK-----	600
QY	728	MAKQAOYVAQLHVWARREHPTYEDIVGLERLYAENPSARDQILPVYAEYQORSEKVLTEY	787
Db	601	-----GAKALAE--SIRPLVAP-----	616
QY	788	NKKALEPVNPKPPPAIKIDPPPPPPQOGLIPGFLMPPSDGSGVTEGTGMPAAAPMVPT	847
Db	617	-----PPAPAPA-PAEPAPAPAPAGEV-----AP-TPTT	643
QY	848	GSPGGGLPA	856
Db	644	PTPQRTLEA	652

Search completed: November 21, 2003, 16:03:20
Job time : 63.515 secs

This Page Blank (uspto)

Db 61 KLEVSFVRPAQPR-GSKPSPGSPETGAGAGTATTTPASSPVTLAETGSLTLYPLFNLWG 119
Qy 121 PAFHERYPNVTITTAQGTSGAGTAQAAAGTNNIGASDAYLSEGDMAHKLGMNIALAISA 180
Db 120 PAFHERYPNVTITTAQGTSGAGTAQAAAGTNNIGASDAYLSEGDMAHKLGMNIALAISA 179
Qy 181 QOVNYPGVSEHLKNGKVLAAVYOGTITKTDWDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QOVNYPGVSEHLKNGKVLAAVYOGTITKTDWDPQIAALNPGVNLPGTAVVPLHRS DGS 239
Qy 241 GDTFLFTQYLSKQDPEGKSGSPGFTTVDPPFVPGALGENGNGMTGCAETPCGVAYIG 300
Db 240 GDTFLFTQYLSKQDPEGKSGSPGFTTVDPPFVPGALGENGNGMTGCAETPCGVAYIG 299
Qy 301 ISFLDQASQRLGEAQLGNSGNNFLPDAQSIQAAAAGFASKTPANQAI SMIDGPADGY 360
Db 300 ISFLDQASQRLGEAQLGNSGNNFLPDAQSIQAAAAGFASKTPANQAI SMIDGPADGY 359
Qy 361 PIINYEYAI VNNRQKDAATQATQAFLEHWAITDGNKASFLDQVHFQPLPVA VVKLS DALI 420
Db 360 PIINYEYAI VNNRQKDAATQATQAFLEHWAITDGNKASFLDQVHFQPLPVA VVKLS DALI 419
Qy 421 ATISSAEMKTDATLAQEAQNFERRISGDLKTQIDQVESTAGSLQGGWRGAAGTAAQA VV 480
Db 420 ATISSAEMKTDATLAQEAQNFERRISGDLKTQIDQVESTAGSLQGGWRGAAGTAAQA VV 479
Qy 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADSEBQQQALSSQMGFV--PTTAA SPPSTAAA 540
Db 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADSEBQQQALSSQMGFV--PTTAA SPPSTAAA 537
Qy 541 ANEVEAPMADPPTDVTTPCELTAANKAAQVLSDNMREYLAAGAKERQRLATSLRNA 600
Db 538 PPAPATVAPPPAAAANTPNAQCGDNAPPPADPNAPPPVPIAPNAPQVR----- 589
Qy 601 AKAYGEVDEEATALNDGSGTVOASAGAVGDS-----SALETDTTPRVATAGEPNF- 653
Db 590 -----IDNPVGGSFALPAGWVESDAAHFDYGSALLS-----KTTGDPPPP 630
Qy 654 -----MDLKEARKELETGQOGASLAHAFADGWNFTNLTLOGGVKBPREGFD 697
Db 631 GQPPPVANDTRIVIGRJDQKLYASAEATDSKAAA-----RLGSDMGEP--YM 675
Qy 698 NWEGDAATACEASLDQORQWILHMAKLSAAMAKAQYVAQLHWARREHPTVEDIVGLER 757
Db 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSKPNQCIWTGVIGSPA 724
Qy 758 LYAENPSARDQIILPVYAQYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
Db 725 ANAPDAGPPORFVFW-----LGTANNPVDKGAAKALAESIRPLVAPPPA-----P 770
Qy 810 PPPPQEQGLIPGLMPPSDGSGVTPTGMPAAPMPVPTGSPGGPLPA 856
Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTQRTLPA 802

RESULT 2

US-09-072-596-209

; Sequence 209, Application US/09072596

; Patent No. 6458366

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, David C.

; APPLICANT: Campos-Neto, Antonia

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedwick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

; NUMBER OF SEQUENCES: 350

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/072,596
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-209

Query Match 53.2%; Score 2700; DB 4; Length 802;

Best Local Similarity 66.68; Pred. No. 5e-187;

Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

Qy 1 NGHHHHHHVDDIIIGTSPTSWEQAAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYI 60
Db 1 NGHHHHHHVDDIIIGTSPTSWEQAAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYI 60
Qy 61 KLEVSFVRPAQPRGSKPSPGSPETGAGAGTATTTPASSPVTLAETGSLTLYPLFNLWG 120
Db 61 KLEVSFVRPAQPRGSKPSPGSPETGAGAGTATTTPASSPVTLAETGSLTLYPLFNLWG 119
Qy 121 PAFHERYPNVTITTAQGTSGAGTAQAAAGTNNIGASDAYLSEGDMAHKLGMNIALAISA 180
Db 120 PAFHERYPNVTITTAQGTSGAGTAQAAAGTNNIGASDAYLSEGDMAHKLGMNIALAISA 179
Qy 181 QOVNYPGVSEHLKNGKVLAAVYOGTITKTDWDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QOVNYPGVSEHLKNGKVLAAVYOGTITKTDWDPQIAALNPGVNLPGTAVVPLHRS DGS 239
Qy 241 GDTFLFTQYLSKQDPEGKSGSPGFTTVDPPFVPGALGENGNGMTGCAETPCGVAYIG 300
Db 240 GDTFLFTQYLSKQDPEGKSGSPGFTTVDPPFVPGALGENGNGMTGCAETPCGVAYIG 299
Qy 301 ISFLDQASQRLGEAQLGNSGNNFLPDAQSIQAAAAGFASKTPANQAI SMIDGPADGY 360
Db 300 ISFLDQASQRLGEAQLGNSGNNFLPDAQSIQAAAAGFASKTPANQAI SMIDGPADGY 359
Qy 361 PIINYEYAI VNNRQKDAATQATQAFLEHWAITDGNKASFLDQVHFQPLPVA VVKLS DALI 420
Db 360 PIINYEYAI VNNRQKDAATQATQAFLEHWAITDGNKASFLDQVHFQPLPVA VVKLS DALI 419
Qy 421 ATISSAEMKTDATLAQEAQNFERRISGDLKTQIDQVESTAGSLQGGWRGAAGTAAQA VV 480
Db 420 ATISSAEMKTDATLAQEAQNFERRISGDLKTQIDQVESTAGSLQGGWRGAAGTAAQA VV 479
Qy 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADSEBQQQALSSQMGFV--PTTAA SPPSTAAA 540
Db 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADSEBQQQALSSQMGFV--PTTAA SPPSTAAA 537
Qy 541 ANEVEAPMADPPTDVTTPCELTAANKAAQVLSDNMREYLAAGAKERQRLATSLRNA 600
Db 538 PPAPATVAPPPAAAANTPNAQCGDNAPPPADPNAPPPVPIAPNAPQVR----- 589

QY 601 AKAYGEVDEBAATDNDGEGTVOAESAGAVGDS-----SALTDTPRVATAGEPNP- 653
DB 590 -----IDNPVGGSFALPAGWVESAAAHFDYGSALLS-----KITGDPFPF 630
QY 654 -----MDLKEARKLETGDCGASLAHFADGWNFTNLTLQGDVKKRFRGFD 697
DB 631 GQPPPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
QY 698 NWEGBAATACBASLDQQRQWILHMAKLSAAMAKQAOYVAQLHWARREHPTIEDIVGLER 757
DB 676 PYPGTRINQETVSLD-----ANGVSGSASYVEVKFSDSPKNGQIWTGVIGSPA 724
QY 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
DB 725 ANAPDAGPPQRFVWV-----LGTANNPVDKGAALAEIRPLVAPPPA-----P 770
QY 810 PPPPQOGLIPGLMPPSDGSGVTPGTGMPAAPMVPPTGSPGGGLPA 856
DB 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 3

US-09-072-596-346
; Sequence 346, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-596-346

Query Match 53.2%; Score 2700; DB 4; Length 802;
Best Local Similarity 66.6%; Pred. No. 5e-187;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

QY 1 MGHHHHHHVIDIGTSPTSWBOAABAVQARSDSDDDIRVARVIBODMAVDGAGKITVYRI 60
DB 1 MGHHHHHHVIDIGTSPTSWEOAABAVQARSDSDDDIRVARVIBODMAVDGAGKITVYRI 60
QY 61 KLEVSFKMPAPRCOSKPPSGSPETGAGAGVATTPASSPVTLAETGTLTLYPLFNLMG 120
DB 61 KLEVSFKMPAPR-GSKPPSGSPETGAGAGVATTPASSPVTLAETGTLTLYPLFNLMG 119
QY 121 PAFHERYPNVTTAQTGSGAGIAQAAAGTNGIGASDAYLSEGDMAAHKGLNIALAISA 180
DB 120 PAFHERYPNVTTAQTGSGAGIAQAAAGTNGIGASDAYLSEGDMAAHKGLNIALAISA 179
QY 181 QQVNYNLPVSHKLNGKVLAAWYQGTIKTWDWQIQAALNPGVNLPGTAVVPLHRSDGS 240
DB 180 QQVNYNLPVSHKLNGKVLAAWYQGTIKTWDWQIQAALNPGVNLPGTAVVPLHRSDGS 239
QY 241 GDTFLETOYLSKQDPPEGNGKSPFGTTPVFPVAPGALGENGNGMVTGCAETPGCVAYIG 300
DB 240 GDTFLETOYLSKQDPPEGNGKSPFGTTPVFPVAPGALGENGNGMVTGCAETPGCVAYIG 299
QY 301 ISFLDOASORGLEAQLGNSGNFLLPDAQSTOAAAANGFASKTPANCAISMIDGRAPDGY 360
DB 300 ISFLDOASORGLEAQLGNSGNFLLPDAQSTOAAAANGFASKTPANCAISMIDGRAPDGY 359
QY 361 PIINYEYAIVNNRQKDAATAQTLQAEFLHWAITDGNKASFLOVHFQFLPPAVVVKLSDALI 420
DB 360 PIINYEYAIVNNRQKDAATAQTLQAEFLHWAITDGNKASFLOVHFQFLPPAVVVKLSDALI 419
QY 421 ATISSAEMKTDATLAQEAQNTFERISGDLKTQIDQVESTAGSLQGWGAAAGTAAQAAV 480
DB 420 ATISSAEMKTDATLAQEAQNTFERISGDLKTQIDQVESTAGSLQGWGAAAGTAAQAAV 479
QY 481 RFOERANKOKOELDEISTNIROAGVOYSRADBEQQOALSQMGFTQSVTVVDQOEILNR 540
DB 480 RFOERANKOKOELDEISTNIROAGVOYSRADBEQQOALSQMGFTQSVTVVDQOEILNR 537
QY 541 ANEVEAPMADPPTDPTTCELTAAKNAQOLVLSADNMREYLAAGAKERQRLATSLRNA 600
DB 538 PPAPATPVAPPVPPAAANTENAPQDENAAPPADPNAPPPVPIAPNAPQVPR----- 589
QY 601 AKAYGEVDEBAATDNDGEGTVOAESAGAVGDS-----SALTDTPRVATAGEPNP- 653
DB 590 -----IDNPVGGSFALPAGWVESAAAHFDYGSALLS-----KITGDPFPF 630
QY 654 -----MDLKEARKLETGDCGASLAHFADGWNFTNLTLQGDVKKRFRGFD 697
DB 631 GQPPPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
QY 698 NWEGBAATACBASLDQQRQWILHMAKLSAAMAKQAOYVAQLHWARREHPTIEDIVGLER 757
DB 676 PYPGTRINQETVSLD-----ANGVSGSASYVEVKFSDSPKNGQIWTGVIGSPA 724
QY 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
DB 725 ANAPDAGPPQRFVWV-----LGTANNPVDKGAALAEIRPLVAPPPA-----P 770
QY 810 PPPPQOGLIPGLMPPSDGSGVTPGTGMPAAPMVPPTGSPGGGLPA 856
DB 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 4

US-09-056-556-184
; Sequence 184, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 184:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-056-556-184

Query Match 46.8%; Score 2375; DB 4; Length 460;

Best Local Similarity 100.0%; Pred. No. 7.3e-164; Mismatches 0; Indels 0; Gaps 0;

QY 525 TQSTVTVDQOEILNRANEVEAPMDPTDPTTCELTAANKAAQQLVLSADNNREYLA 584
DB 2 TQSTVTVDQOEILNRANEVEAPMDPTDPTTCELTAANKAAQQLVLSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAKAYGEVDEEAATALDNDGEGTVOAESAGAVGGDSSAELTDTTPR 644
DB 62 AGAKERQRLATSLRNAKAYGEVDEEAATALDNDGEGTVOAESAGAVGGDSSAELTDTTPR 121
QY 645 VATAGEPNFMDLKEAARKLETGQGLASLAHFAFGWNTFNLTLQGDVXKFRGFDNNEGDA 704
DB 122 VATAGEPNFMDLKEAARKLETGQGLASLAHFAFGWNTFNLTLQGDVXKFRGFDNNEGDA 181
QY 705 TACEASLDQORQWILHMAKLSAAMAKQAQYVAQLHWARREHPTTYEDIVGLERLYAENPS 764
DB 182 TACEASLDQORQWILHMAKLSAAMAKQAQYVAQLHWARREHPTTYEDIVGLERLYAENPS 241
QY 765 ARDQILPVYAEYQORSEKVLTEYNKAALEPVNPPKPPAIKIDPPPPQEOGLIPGFLM 824
DB 242 ARDQILPVYAEYQORSEKVLTEYNKAALEPVNPPKPPAIKIDPPPPQEOGLIPGFLM 301
QY 825 PPSDGSVTPGTGMPAAPVPPPTGSGGGLPADTAQAQLTSAGREAAALSGDVAVKAAISG 884
DB 302 PPSDGSVTPGTGMPAAPVPPPTGSGGGLPADTAQAQLTSAGREAAALSGDVAVKAAISG 361
QY 885 GGGGGVPSAPLGSALGGAESVVRPACAGDIAGLQGRAGGGAALGGGGMGMPGAAHQOQ 944
DB 362 GGGGGVPSAPLGSALGGAESVVRPACAGDIAGLQGRAGGGAALGGGGMGMPGAAHQOQ 421
QY 945 GGAKSQSGQDEALYTEDRAWTEAVIGNRRQDSKESK 983
DB 422 GGAKSQSGQDEALYTEDRAWTEAVIGNRRQDSKESK 460

RESULT 5

US-09-072-596-179
; Sequence 179, Application US/09072596
; Patent No. 6459366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 179:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-179

Query Match 46.8%; Score 2375; DB 4; Length 460;

Best Local Similarity 100.0%; Pred. No. 7.3e-164; Mismatches 0; Indels 0; Gaps 0;

QY 525 TQSTVTVDQOEILNRANEVEAPMDPTDPTTCELTAANKAAQQLVLSADNNREYLA 584
DB 2 TQSTVTVDQOEILNRANEVEAPMDPTDPTTCELTAANKAAQQLVLSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAKAYGEVDEEAATALDNDGEGTVOAESAGAVGGDSSAELTDTTPR 644
DB 62 AGAKERQRLATSLRNAKAYGEVDEEAATALDNDGEGTVOAESAGAVGGDSSAELTDTTPR 121
QY 645 VATAGEPNFMDLKEAARKLETGQGLASLAHFAFGWNTFNLTLQGDVXKFRGFDNNEGDA 704
DB 122 VATAGEPNFMDLKEAARKLETGQGLASLAHFAFGWNTFNLTLQGDVXKFRGFDNNEGDA 181
QY 705 TACEASLDQORQWILHMAKLSAAMAKQAQYVAQLHWARREHPTTYEDIVGLERLYAENPS 764
DB 182 TACEASLDQORQWILHMAKLSAAMAKQAQYVAQLHWARREHPTTYEDIVGLERLYAENPS 241
QY 765 ARDQILPVYAEYQORSEKVLTEYNKAALEPVNPPKPPAIKIDPPPPQEOGLIPGFLM 824
DB 242 ARDQILPVYAEYQORSEKVLTEYNKAALEPVNPPKPPAIKIDPPPPQEOGLIPGFLM 301
QY 825 PPSDGSVTPGTGMPAAPVPPPTGSGGGLPADTAQAQLTSAGREAAALSGDVAVKAAISG 884
DB 302 PPSDGSVTPGTGMPAAPVPPPTGSGGGLPADTAQAQLTSAGREAAALSGDVAVKAAISG 361
QY 885 GGGGGVPSAPLGSALGGAESVVRPACAGDIAGLQGRAGGGAALGGGGMGMPGAAHQOQ 944
DB 362 GGGGGVPSAPLGSALGGAESVVRPACAGDIAGLQGRAGGGAALGGGGMGMPGAAHQOQ 421

QY 945 GGAAGKSGQDEALYTEDRAWTEAVIGNRRQDSKSK 983
DB 422 GGAAGKSGQDEALYTEDRAWTEAVIGNRRQDSKSK 460

RESULT 6

US-09-072-596-350
Sequence 350, Application US/09072596
Patent No. 6458366

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.

TUBERCULOSIS

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 350:

SEQUENCE CHARACTERISTICS:

LENGTH: 652 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-072-596-350.

Query Match 36.3%; Score 1839; DB 4; Length 652;
Best Local Similarity 48.5%; Pred. No. 7, 7e-125;
Matches 441; Conservative 26; Mismatches 133; Indels 310; Gaps 19;

QY 1 MGHHHHHHVILIGTSWEQAAAEAVQARADSVDDIRVARVIEQDMVDSAGKITVRI 60

DB 1 MGHHHHHH----- 8

QY 61 KLEVSFKRPAQPRCKSPGSPETGAGAGTVATTPASSPVTLAETGTLTYPLFNLWG 120

DB 9 -----GSPKPSGSPETGAGAGTVATTPASSPVTLAETGTLTYPLFNLWG 53

QY 121 PAFHERYPNVITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNLALISA 180

DB 54 PAFHERYPNVITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNLALISA 113

QY 181 QGVNYNLPGVSEHLKLVLAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDGS 240

DB 114 QGVNYNLPGVSEHLKLVLAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDGS 173

QY 241 GDTFLFTQYLSKODPEGWKGSPGFTTVDPPPAVPGALGENGNGMVTGCAETPGCVAYIG 300
DB 174 GDTFLFTQYLSKODPEGWKGSPGFTTVDPPPAVPGALGENGNGMVTGCAETPGCVAYIG 233
QY 301 ISFLDQASQSGGLGEAQLGNSGNGNLLPDAQSIQAAAAAGFASKTPANQAISMIDGPAPDGY 360
DB 234 ISFLDQASQSGGLGEAQLGNSGNGNLLPDAQSIQAAAAAGFASKTPANQAISMIDGPAPDGY 293
QY 361 PIINYEYAIYNNRQKDAATAQTLQAFHWAITDGNKASFLDQVHFQPLPPAVVVKLSDALI 420
DB 294 PIINYEYAIYNNRQKDAATAQTLQAFHWAITDGNKASFLDQVHFQPLPPAVVVKLSDALI 353
QY 421 ATISSAEMKTDAAATLACEAGNPERISGDLKTQIDQVESTAGSLQGWRAAGTAQAQAAVV 480
DB 354 ATISSGG-----GSGGGSGGSGSVPTTAASPPSTA-----AAPAPATPVA 396
QY 481 RFOEAANKQKQELDEISTNIRQAGVOYVSRADBEQQOALSSQMGFTQSTQTVTDVQOBILNR 540
DB 397 PPPPAA-----ANTPNAQPG----- 411
QY 541 ANEVEAPMADPPPTDVPITPCELTAAKNAACQLVLSADN---MREYLAAGAKERQ----- 591
DB 412 -----DPNAAPPADPNAPPPVIAPNAPQPVRI--DNVGGSPFALPAGWVESDAHFD 454
QY 592 -----RLATSLRNAAKAYGEVDEEAATALDND-- 618
DB 465 YGSALLSKTTGDPFPFPGQPPPVANDTRIVLGRLDQKLYASAEA---TDSKAAARLGSMDMG 521
QY 619 -----GEGTQVQASAGAVGSDSAAELTDTTPRVATAGEPNMDLKEAARKLETGD 667
DB 522 EYMPYVPGTRINQETVSLDANGVSSGSAAYEV-----KFSDFSPKNGQIWTGV 569
QY 668 QGASLAHFADGWNFTNLTLQGDVFRFGDNWEGDAATACEASLDQORQWILHMAKLISAA 727
DB 570 IGSPAANAPDA-----GPFQRW--FVWLGTANNPVDK----- 600
QY 728 MAKQAOYVLAQLHVWARREHPTVEDIVGLERLYAENSPSARDQILPVYAEVQORSEKVLTEY 787
DB 601 -----GAAKALAE--SIRPLVAP----- 616
QY 788 NKAALPEVNNPKPPPAIKIDPPPPQEQGLIFGLMPPSDGSGVTGTCMPAAPVPPPT 847
DB 617 -----PPAPAPA-PAEPAPAPAPAGEV-----AP-TPPT 643
QY 848 GSPFGGLPA 856
DB 644 PTFORTLPA 652

RESULT 7

US-08-818-112-153

Sequence 153, Application US/08818112

Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, David C.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Vedvick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/618,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-618-112-153

Query Match 36.1%; Score 1831; DB 3; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;
QY 59 RIKLEVSFKMRPAQ-----RCGSKPPSGSPETGAGAGTVAATTPASSPVTLAETGSTALL 112
DB 2 KIRHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGAGTVAATTPASSPVTLAETGSTALL 61
QY 113 YPLFNLWGPAFHRYPNVITTAQGTGSGAGIAQAAGTGNIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLFNLWGPAFHRYPNVITTAQGTGSGAGIAQAAGTGNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAAQVYNLPGVSEHLKXNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAAQVYNLPGVSEHLKXNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDSGDTFTFTQYLSKQDPEGKSGFPGFTTVDFFPAVPGALGNGNGMVTGCAET 292
DB 182 PLHRSDSGDTFTFTQYLSKQDPEGKSGFPGFTTVDFFPAVPGALGNGNGMVTGCAET 241
QY 293 PGCVAYIGISFLDQASQRLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTTPANQAISMI 352
DB 242 PGCVAYIGISFLDQASQRLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTTPANQAISMI 301
QY 353 DGPAPDGYPIINYEYAIYVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAIYVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

RESULT 8
US-08-618-111-148
Sequence 148, Application US/08618111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/618,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-618-111-148

Query Match 36.1%; Score 1831; DB 4; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;
QY 59 RIKLEVSFKMRPAQ-----RCGSKPPSGSPETGAGAGTVAATTPASSPVTLAETGSTALL 112
DB 2 KIRHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGAGTVAATTPASSPVTLAETGSTALL 61
QY 113 YPLFNLWGPAFHRYPNVITTAQGTGSGAGIAQAAGTGNIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLFNLWGPAFHRYPNVITTAQGTGSGAGIAQAAGTGNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAAQVYNLPGVSEHLKXNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAAQVYNLPGVSEHLKXNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDSGDTFTFTQYLSKQDPEGKSGFPGFTTVDFFPAVPGALGNGNGMVTGCAET 292
DB 182 PLHRSDSGDTFTFTQYLSKQDPEGKSGFPGFTTVDFFPAVPGALGNGNGMVTGCAET 241
QY 293 PGCVAYIGISFLDQASQRLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTTPANQAISMI 352
DB 242 PGCVAYIGISFLDQASQRLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTTPANQAISMI 301
QY 353 DGPAPDGYPIINYEYAIYVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAIYVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

RESULT 9
US-09-056-556-153
Sequence 153, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue

```

; CITY: ~ Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-056-556-153

Query Match 36.1%; Score 1831; DB 4; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKRPAQP-----RCGSKPPSGSPETGAGAGTATTTPASSPVTTLAETGSTALL 112
Db 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGAGTATTTPASSPVTTLAETGSTALL 61
QY 113 YPLENLWGFAPHERYPNVTTTAQCTGSGAGIAQAAAGTNNICASDAYLSEGDMAAHKGLM 172
Db 62 YPLENLWGFAPHERYPNVTTTAQCTGSGAGIAQAAAGTNNICASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAQVNNYLPVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVV 232
Db 122 NIALAISAQVNNYLPVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDSGDTFTLYQLSKQDPEGWKGSPGFGTTVDPAVFGALGNGNGMVGCAET 292
Db 182 PLHRSDSGDTFTLYQLSKQDPEGWKGSPGFGTTVDPAVFGALGNGNGMVGCAET 241
QY 293 PCVAYIGISFLDQASQGLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMI 352
Db 242 PCVAYIGISFLDQASQGLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMI 301
QY 353 DGPAPDGYPIINVEYAI VNNRQKDAATAQTLOAFLHWAITDGNKASFLLDQVHFQPLPPAV 412
Db 302 DGPAPDGYPIINVEYAI VNNRQKDAATAQTLOAFLHWAITDGNKASFLLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
Db 362 VKLSDALIATISS 374

```

```

RESULT 10
US-09-056-556-155
; Sequence 155, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP

```

```

; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 155:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-056-556-155

Query Match 36.1%; Score 1831; DB 4; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKRPAQP-----RCGSKPPSGSPETGAGAGTATTTPASSPVTTLAETGSTALL 112
Db 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGAGTATTTPASSPVTTLAETGSTALL 61
QY 113 YPLENLWGFAPHERYPNVTTTAQCTGSGAGIAQAAAGTNNICASDAYLSEGDMAAHKGLM 172
Db 62 YPLENLWGFAPHERYPNVTTTAQCTGSGAGIAQAAAGTNNICASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAQVNNYLPVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVV 232
Db 122 NIALAISAQVNNYLPVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDSGDTFTLYQLSKQDPEGWKGSPGFGTTVDPAVFGALGNGNGMVGCAET 292
Db 182 PLHRSDSGDTFTLYQLSKQDPEGWKGSPGFGTTVDPAVFGALGNGNGMVGCAET 241
QY 293 PCVAYIGISFLDQASQGLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMI 352
Db 242 PCVAYIGISFLDQASQGLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMI 301
QY 353 DGPAPDGYPIINVEYAI VNNRQKDAATAQTLOAFLHWAITDGNKASFLLDQVHFQPLPPAV 412
Db 302 DGPAPDGYPIINVEYAI VNNRQKDAATAQTLOAFLHWAITDGNKASFLLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
Db 362 VKLSDALIATISS 374

RESULT 11
US-09-072-596-148
; Sequence 148, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.

```

```

; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-596-148

```

```

Query Match 36.1%; Score 1831; DB 4; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKVRPAQP-----RCGSKPPSGSPETGAGTVAATTPASSPVTLAETGSTALL 112
DB 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGTVAATTPASSPVTLAETGSTALL 61

QY 113 YPLENLGPAFHRYPNVTTTAQGTSGAGIAQAAGTNIIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLENLGPAFHRYPNVTTTAQGTSGAGIAQAAGTNIIGASDAYLSEGDMAAHKGLM 121

QY 173 NIALAISAOQVYNLPGVSEHLKNGKVLAAAYQGTTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAOQVYNLPGVSEHLKNGKVLAAAYQGTTIKTWDDPQIAALNPGVNLPGTAVV 181

QY 233 PLHRSDSGDTFLFTQYLSKQDPGEGWKSPFGTGTVDFFAVPGALGNGGNGMTGCAET 292
DB 182 PLHRSDSGDTFLFTQYLSKQDPGEGWKSPFGTGTVDFFAVPGALGNGGNGMTGCAET 241

QY 293 PGCVAYIGISFLDOASORGIGEAOLGNSGNFLLPDAQSIQAAAAGFASKTPANQAISMI 352
DB 242 PGCVAYIGISFLDOASORGIGEAOLGNSGNFLLPDAQSIQAAAAGFASKTPANQAISMI 301

QY 353 DGPAPDGYPIINYEYAIIVNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAIIVNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 361

QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

```

RESULT 12
 US-09-072-596-150
 ; Sequence 150, Application US/09072596
 ; Patent No. 6458366

```

; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 150:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-596-150

```

```

Query Match 36.1%; Score 1831; DB 4; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKVRPAQP-----RCGSKPPSGSPETGAGTVAATTPASSPVTLAETGSTALL 112
DB 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGTVAATTPASSPVTLAETGSTALL 61

QY 113 YPLENLGPAFHRYPNVTTTAQGTSGAGIAQAAGTNIIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLENLGPAFHRYPNVTTTAQGTSGAGIAQAAGTNIIGASDAYLSEGDMAAHKGLM 121

QY 173 NIALAISAOQVYNLPGVSEHLKNGKVLAAAYQGTTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAOQVYNLPGVSEHLKNGKVLAAAYQGTTIKTWDDPQIAALNPGVNLPGTAVV 181

QY 233 PLHRSDSGDTFLFTQYLSKQDPGEGWKSPFGTGTVDFFAVPGALGNGGNGMTGCAET 292
DB 182 PLHRSDSGDTFLFTQYLSKQDPGEGWKSPFGTGTVDFFAVPGALGNGGNGMTGCAET 241

QY 293 PGCVAYIGISFLDOASORGIGEAOLGNSGNFLLPDAQSIQAAAAGFASKTPANQAISMI 352
DB 242 PGCVAYIGISFLDOASORGIGEAOLGNSGNFLLPDAQSIQAAAAGFASKTPANQAISMI 301

QY 353 DGPAPDGYPIINYEYAIIVNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAIIVNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 361

QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

```

Db 362 VKLSDALIATISS 374

RESULT 13

US-09-118-426-5

Sequence 5, Application US/09118426C

Patent No. 6517839

GENERAL INFORMATION:

APPLICANT: Modlin, Robert L.

APPLICANT: Libraty, Daniel H.

TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

TITLE OF INVENTION: T-CELL RESPONSE

FILE REFERENCE: 30435.4US01

CURRENT APPLICATION NUMBER: US/09/118,426C

CURRENT FILING DATE: 1998-07-17

EARLIER APPLICATION NUMBER: 60/052,970

EARLIER FILING DATE: 1998-07-17

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 373

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

PUBLICATION INFORMATION:

JOURNAL: Infect. Immun.

VOLUME: 57

ISSUE: 8

PAGES: 2481-

DATE: 1989

US-09-118-426-5

Query Match 36.1%; Score 1829; DB 4; Length 373;

Best Local Similarity 100.0%; Pred. No. 1.8e-124;

Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 CGSKPPSGPETGAGAGTATTTPASSPVTLAETGTLTLLYPLNMGPAFHRYPNVTITA 134

Db 23 CGSKPPSGPETGAGAGTATTTPASSPVTLAETGTLTLLYPLNMGPAFHRYPNVTITA 82

QY 135 QGTGSGAGTAQAAGTNTGASDAYLSEGDMAHKGMLNIALAISAOQVNNLPGVSEHL 194

Db 83 QGTGSGAGTAQAAGTNTGASDAYLSEGDMAHKGMLNIALAISAOQVNNLPGVSEHL 142

QY 195 KLGKVLAAMYOGTITKTDWDDPQIAALNPGVNLPGTAVVPLHRSKQD 254

Db 143 KLGKVLAAMYOGTITKTDWDDPQIAALNPGVNLPGTAVVPLHRSKQD 202

QY 255 PEGWKSFGTITVDPFPAVPGALGENGGMTGCAETPGCVAYIGISFLDQASQRLGE 314

Db 203 PEGWKSFGTITVDPFPAVPGALGENGGMTGCAETPGCVAYIGISFLDQASQRLGE 262

QY 315 AOLGNSSGNFLPDQAQSIQAAAAGFASKTPANQAISMIDGPDGYPPIINYEYAVNNRQ 374

Db 263 AOLGNSSGNFLPDQAQSIQAAAAGFASKTPANQAISMIDGPDGYPPIINYEYAVNNRQ 322

QY 375 KDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKLSDALIATISS 425

Db 323 KDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKLSDALIATISS 373

RESULT 14

US-09-118-426-6

Sequence 6, Application US/09118426C

Patent No. 6517839

GENERAL INFORMATION:

APPLICANT: Modlin, Robert L.

APPLICANT: Libraty, Daniel H.

TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

TITLE OF INVENTION: T-CELL RESPONSE

FILE REFERENCE: 30435.4US01

CURRENT APPLICATION NUMBER: US/09/118,426C

CURRENT FILING DATE: 1998-07-17

EARLIER APPLICATION NUMBER: 60/052,970

US-09-118-426-6

Query Match 35.9%; Score 1820; DB 4; Length 351;

Best Local Similarity 100.0%; Pred. No. 7.4e-124;

Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 GSKPPSGPETGAGAGTATTTPASSPVTLAETGTLTLLYPLNMGPAFHRYPNVTITA 135

Db 2 GSKPPSGPETGAGAGTATTTPASSPVTLAETGTLTLLYPLNMGPAFHRYPNVTITA 61

QY 136 GTGSGAGTAQAAGTNTGASDAYLSEGDMAHKGMLNIALAISAOQVNNLPGVSEHLX 195

Db 62 GTGSGAGTAQAAGTNTGASDAYLSEGDMAHKGMLNIALAISAOQVNNLPGVSEHLX 121

QY 196 LNKVLAAMYOGTITKTDWDDPQIAALNPGVNLPGTAVVPLHRSKQD 255

Db 122 LNKVLAAMYOGTITKTDWDDPQIAALNPGVNLPGTAVVPLHRSKQD 181

QY 256 EGWKSFGTITVDPFPAVPGALGENGGMTGCAETPGCVAYIGISFLDQASQRLGE 315

Db 182 EGWKSFGTITVDPFPAVPGALGENGGMTGCAETPGCVAYIGISFLDQASQRLGE 241

QY 316 QLGNSGNFLPDQAQSIQAAAAGFASKTPANQAISMIDGPDGYPPIINYEYAVNNRQ 375

Db 242 QLGNSGNFLPDQAQSIQAAAAGFASKTPANQAISMIDGPDGYPPIINYEYAVNNRQ 301

QY 376 DAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKLSDALIATISS 425

Db 302 DAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKLSDALIATISS 351

RESULT 15

US-08-818-112-89

Sequence 89, Application US/08818112

Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Vedwick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,112

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Mark David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-89

Query Match      15.6%; Score 791; DB 3; Length 166;
Best Local Similarity 95.2%; Pred. No. 7.4e-50;
Matches 157; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 525 TQQTIVTDQOEILNRANEVEAPMDPTDVPITPCELTAAKNAQQLVLSADNMREYLA 584
Db 2 TQQTIVTDQOEILNRANEVEAPMDPTDVPITPCELTXKNAQQLVLSADNMREYLA 61

Qy 595 AGAKERORLATSRLNAKAYGEVDEEAATALDNDGEGTVQAESAGAVGGSSAELTDTTPR 644
Db 62 AGAKERORLATSRLNAKAYGEVDEEAATALDNDGEGTVQAESAGAVGGSSAELTDTTPR 121

Qy 645 VATAGEFNFMDLKEAAKLETGQGLAHFADGWNFTNLTQGD 689
Db 122 VATAGEFNFMDLKEAAKLETGQGLAHFADGWNFTNLTQGD 166

Search completed: November 21, 2003, 16:11:43
Job time : 22.7299 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 38.3392 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-54

Perfect score: 5072

Sequence: 1 MGHHHHHVIDIIGTSPTW.....RANTEAVIGNRRQDSKESK 983

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PTCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PTCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2700	53.2	802	9	US-09-287-849-10
2	2700	53.2	802	12	US-10-084-843-214
3	2700	53.2	802	12	US-10-084-843-351
4	2700	53.2	802	12	US-10-084-843-351
5	2700	53.2	802	12	US-10-193-002-209
6	2700	53.2	802	12	US-10-193-002-346
7	2700	53.2	802	12	US-10-193-002-346
8	2700	53.2	802	12	US-10-193-002-346
9	2700	53.2	802	12	US-10-193-002-346
10	2700	53.2	802	12	US-10-193-002-346
11	2700	53.2	802	12	US-10-193-002-346
12	2700	53.2	802	12	US-10-193-002-346
13	2700	53.2	802	12	US-10-193-002-346
14	2700	53.2	802	12	US-10-193-002-346
15	2700	53.2	802	12	US-10-193-002-346

```
16 1831 36.1 374 12 US-10-193-002-150 Sequence 150, Appl
17 1831 36.1 374 12 US-10-359-460-6 Sequence 6, Appl
18 1831 36.1 374 12 US-10-359-460-40 Sequence 40, Appl
19 1831 36.1 374 12 US-10-098-732A-39 Sequence 39, Appl
20 791 15.6 166 12 US-10-084-843-89 Sequence 89, Appl
21 791 15.6 166 12 US-10-193-002-90 Sequence 90, Appl
22 485 9.6 100 12 US-10-084-843-115 Sequence 115, Appl
23 485 9.6 100 12 US-10-193-002-110 Sequence 110, Appl
24 485 9.6 100 16 US-10-080-170-639 Sequence 639, Appl
25 481 9.5 100 15 US-10-140-045-5 Sequence 5, Appl
26 479.5 9.5 358 9 US-09-287-849-8 Sequence 8, Appl
27 479.5 9.5 358 12 US-10-359-460-8 Sequence 8, Appl
28 466 9.2 95 12 US-10-084-843-88 Sequence 88, Appl
29 466 9.2 95 12 US-10-193-002-89 Sequence 89, Appl
30 466 9.2 95 12 US-10-098-732A-35 Sequence 35, Appl
31 413 8.1 346 9 US-09-741-669-380 Sequence 380, Appl
32 392 7.7 80 12 US-10-084-843-117 Sequence 117, Appl
33 392 7.7 80 12 US-10-193-002-112 Sequence 112, Appl
34 329.5 6.5 364 12 US-10-084-843-73 Sequence 73, Appl
35 329.5 6.5 364 12 US-10-193-002-74 Sequence 74, Appl
36 324 6.4 66 12 US-10-084-843-77 Sequence 77, Appl
37 324 6.4 66 12 US-10-193-002-78 Sequence 78, Appl
38 324 6.4 66 12 US-10-098-732A-37 Sequence 37, Appl
39 313.5 6.2 368 15 US-10-156-761-11606 Sequence 11606, A
40 276 5.4 375 10 US-09-738-626-6333 Sequence 6333, Ap
41 236 4.7 49 15 US-10-140-045-6 Sequence 6, Appl
42 217.5 4.3 1463 10 US-09-971-536-69 Sequence 69, Appl
43 205 4.0 42 15 US-10-140-045-7 Sequence 7, Appl
44 202.5 4.0 2468 12 US-10-246-330-4 Sequence 4, Appl
45 192 3.8 856 9 US-09-287-849-12 Sequence 12, Appl
```

ALIGNMENTS

RESULT 1

```
US-09-287-849-10
; Sequence 10, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-008020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-10
```

Query Match 53.2% Score 2700; DB 9; Length 802;
Best Local Similarity 66.6%; Pred. No. 6.1e-161;

Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

Qy 1 MGHVHHVHVDIIGTSTSEWQAAEAQVARSVDVDIRVARVIEQDMVDSAGKITRYI 60
Db 1 MGHVHHVHVDIIGTSTSEWQAAEAQVARSVDVDIRVARVIEQDMVDSAGKITRYI 60

Qy 61 KLEVSFMRPAQPCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGTLTLLPLFNLWG 120
Db 61 KLEVSFMRPAQPR-GSKPPSGSPETGAGAGTATTTPASSPVTLAETGTLTLLPLFNLWG 119

Qy 121 PAFHERYPNVTITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAHKLNMIALAISA 180
Db 120 PAFHERYPNVTITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAHKLNMIALAISA 179

Qy 181 QQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239

Qy 241 GDTFLFTQYLSKODPEGWKSPPGFTTVPFPAVPGALGENGGMVTCGAETPGCVAYIG 300
Db 240 GDTFLFTQYLSKODPEGWKSPPGFTTVPFPAVPGALGENGGMVTCGAETPGCVAYIG 299

Qy 301 ISFLDQASQRLGEAQLGNSSGNFLPLDPAQSIQAAAAGFASKTPANQAISMIDGPADGY 360
Db 300 ISFLDQASQRLGEAQLGNSSGNFLPLDPAQSIQAAAAGFASKTPANQAISMIDGPADGY 359

Qy 361 PIINYEYAI VNNRQKDAATQTLQAFHLWAITDGNKASFLDQVHFQPLPAPVVKLS DALI 420
Db 360 PIINYEYAI VNNRQKDAATQTLQAFHLWAITDGNKASFLDQVHFQPLPAPVVKLS DALI 419

Qy 421 ATTSSBEMKTDATLAQEAQNFRIISGDLTKQIDQVESTAGSLQGWGAAAGTAAQAAV 480
Db 420 ATTSSBEMKTDATLAQEAQNFRIISGDLTKQIDQVESTAGSLQGWGAAAGTAAQAAV 479

Qy 481 RFOEAANKQKQELDEISTNIRQAGVYSRADBEQQQALSSQMGFTQSTVTVDDQEI LNR 540
Db 480 RFOEAANKQKQELDEISTNIRQAGVYSRADBEQQQALSSQMGFTV--PTTASPPSTAAA 537

Qy 541 ANEVEAPMADPTDPTITPCELTAANKAAQQLVLSADNVREYLAAGAKERQRLATSLRNA 600
Db 538 PPAPATPVAPPPPPAAAANTENAPGDPNAPPPPADPNAPPPVIAPNAPQPV-- 589

Qy 601 AKAYGEVDEBAATALNDGEGTVQAESAGAVGDS-----SAELTDITRVATAGEPNF- 653
Db 590 -----INPVGGSFALPAGWVESDAHPDYGSALLS-----KTTGDPFPF 630

Qy 654 -----MDLKEARKLETGDCGASLAHFAADGNNTFNLTILQGVKRFGRFD 697
Db 631 GQPPPVANDTRIVLGRLDQKLYASAEATDSKAAA-----RLGSDWGEF--YM 675

Qy 698 NWEGDAATACEASLDQORQWILHMAKLSAAMAKAQVVAQLHVHWAARREHPTVEDIVGLER 757
Db 676 PYGTTRINQETVSLD-----ANGVSGSASYIEVKFSDPKPNQIQTGVTGSPA 724

Qy 758 LYAENPSARDQILPVVAYEQCRSEKVLTEYNN-----KAALPEVNP--PKPPPAIKIDP 809
Db 725 ANAPDAGPQPRWFWV-----LGTANNFVDKGAALAEIRPLVAPPPA----P 770

Qy 810 PPPPQCGLIPGLMPPSGSGVTPGTGMPAAMPVPPPTGSPCGGLPA 856
Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLEA 802

RESULT 2

US-10-084-843-214
; Sequence 214, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond

Vedvick, Thomas S.
Iwardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ. ID NO: 214:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 214:
US-10-084-843-214

Query Match 53.28; Score 2700; DB 12; Length 802;
Best Local Similarity 66.64; Pred. No. 6.1e-161;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

Qy 1 MGHVHHVHVDIIGTSTSEWQAAEAQVARSVDVDIRVARVIEQDMVDSAGKITRYI 60
Db 1 MGHVHHVHVDIIGTSTSEWQAAEAQVARSVDVDIRVARVIEQDMVDSAGKITRYI 60

Qy 61 KLEVSFMRPAQPCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGTLTLLPLFNLWG 120
Db 61 KLEVSFMRPAQPR-GSKPPSGSPETGAGAGTATTTPASSPVTLAETGTLTLLPLFNLWG 119

Qy 121 PAFHERYPNVTITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAHKLNMIALAISA 180
Db 120 PAFHERYPNVTITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAHKLNMIALAISA 179

Qy 181 QQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239

Qy 241 GDTFLFTQYLSKODPEGWKSPPGFTTVPFPAVPGALGENGGMVTCGAETPGCVAYIG 300
Db 240 GDTFLFTQYLSKODPEGWKSPPGFTTVPFPAVPGALGENGGMVTCGAETPGCVAYIG 299

Qy 301 ISFLDQASQRLGEAQLGNSSGNFLPLDPAQSIQAAAAGFASKTPANQAISMIDGPADGY 360
Db 300 ISFLDQASQRLGEAQLGNSSGNFLPLDPAQSIQAAAAGFASKTPANQAISMIDGPADGY 359

Qy 361 PIINYEYAI VNNRQKDAATQTLQAFHLWAITDGNKASFLDQVHFQPLPAPVVKLS DALI 420
Db 360 PIINYEYAI VNNRQKDAATQTLQAFHLWAITDGNKASFLDQVHFQPLPAPVVKLS DALI 419

421 ATISSAEMKTDATLAQEAAGNFERISGDLTKTQIDQVESTAGSLOGQREGAGTAQAQAVV 480
Db
420 ATISSAEMKTDATLAQEAAGNFERISGDLTKTQIDQVESTAGSLOGQREGAGTAQAQAVV 479
Qy 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGFTQSQTVTVDOQELNR 540
Db 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGFV--PTTAASPPSTAAA 537
Qy 541 ANEVEAPMADPPTDVPITPCELTAAKNAQQLVLSADNMREYLAAGAKERQRLATSLRNA 600
Db 538 PPAPATFVAPPPPPAAANTPNAQCPDNPAPPPADPNAPPPVIAFNAPQFVR----- 589
Qy 601 AKAYGEVDEEAATALNDGEGTVQAESAGAVGDS-----SALTDTPRVATAGBNF- 653
Db 590 -----IDNPVGGFSFALPAGWVESDAAHFDYGSALLS-----KITGDPFPP 630
Qy 654 -----MDLKEAARKLETGDOGASLAHFAWGNTFNLTLOGDVKFRFGFD 697
Db 631 GQPPPVANDTRIVLGRDQKLYASAEATDSKAAA-----RLGSDMGFE--YM 675
Qy 698 NWEGDAATACEASLDQOQWILHMAKLSAAMAKQAQYVAQLHVAWREHPTIEDIVGLER 757
Db 676 PYGTRINQETVSLD-----ANGVSGSASYEVKFSDFSKENGQIWTGVIGSPA 724
Qy 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
Db 725 ANAPDAGPPQORWFFVW-----LGTANNPVDKGAALAESIRPLVAPPPA----P 770
Qy 810 PPPQEOGLIPGFLMPPSDSGVTPGTGMPAPMVPTGSPGGGLPA 856
Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 3
US-10-084-843-351
; Sequence 351, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, David C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; State: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 351:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 351:
US-10-084-843-351

Query Match 53.2%; Score 2700; DB 12; Length 802;
Best Local Similarity 66.6%; Pred. No. 6.1e-161;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

Qy 1 MGHHHHHHVIDIGTSPTSWEQAAAPAVQARSDSDVDIRVARVIEDQMAVDASGKITRYI 60
Db 1 MGHHHHHHVIDIGTSPTSWEQAAAEAVQARSDSDVDIRVARVIEDQMAVDASGKITRYI 60
Qy 61 KLEVSFKMRPAOPRCGSKPPSGSPETGAGAGTATTPASSPVTTLAETGSTLLYPLFNMG 120
Db 61 KLEVSFKMRPAOPR-GSKPPSGSPETGAGAGTATTPASSPVTTLAETGSTLLYPLFNMG 119
Qy 121 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNLIALISA 180
Db 120 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNLIALISA 179
Qy 181 QQVYNLPGVSEHLKLVLAAMVQGTITKTDWDDQIQAALNFGVNLPGTAVPLHRS DGS 240
Db 180 QQVYNLPGVSEHLKLVLAAMVQGTITKTDWDDQIQAALNFGVNLPGTAVPLHRS DGS 239
Qy 241 GDTFLTQVLSKODPEGWCKSPGFGTTVPFPAVPGALGNGNGMVTCGAETPGCVAYIG 300
Db 240 GDTFLTQVLSKODPEGWCKSPGFGTTVPFPAVPGALGNGNGMVTCGAETPGCVAYIG 299
Qy 301 ISFLDQASQRLGEAQGLNNSGNFLLPDAQSIQAAAAGFASKTPANQA-SMIDGRAPDGY 360
Db 300 ISFLDQASQRLGEAQGLNNSGNFLLPDAQSIQAAAAGFASKTPANQA-SMIDGRAPDGY 359
Qy 361 PIINYEAVNNRQKDAATAQTLOAFHLWALTIDGNKASFLDOVHFOPLPVPVVKLSDALI 420
Db 360 PIINYEAVNNRQKDAATAQTLOAFHLWALTIDGNKASFLDOVHFOPLPVPVVKLSDALI 419
Qy 421 ATISSAEMKTDATLAQEAAGNFERISGDLTKTQIDQVESTAGSLOGQREGAGTAQAQAVV 480
Db 420 ATISSAEMKTDATLAQEAAGNFERISGDLTKTQIDQVESTAGSLOGQREGAGTAQAQAVV 479
Qy 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGFTQSQTVTVDOQELNR 540
Db 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGFV--PTTAASPPSTAAA 537
Qy 541 ANEVEAPMADPPTDVPITPCELTAAKNAQQLVLSADNMREYLAAGAKERQRLATSLRNA 600
Db 538 PPAPATFVAPPPPPAAANTPNAQCPDNPAPPPADPNAPPPVIAFNAPQFVR----- 589
Qy 601 AKAYGEVDEEAATALNDGEGTVQAESAGAVGDS-----SALTDTPRVATAGBNF- 653
Db 590 -----IDNPVGGFSFALPAGWVESDAAHFDYGSALLS-----KITGDPFPP 630
Qy 654 -----MDLKEAARKLETGDOGASLAHFAWGNTFNLTLOGDVKFRFGFD 697
Db 631 GQPPPVANDTRIVLGRDQKLYASAEATDSKAAA-----RLGSDMGFE--YM 675
Qy 698 NWEGDAATACEASLDQOQWILHMAKLSAAMAKQAQYVAQLHVAWREHPTIEDIVGLER 757
Db 676 PYGTRINQETVSLD-----ANGVSGSASYEVKFSDFSKENGQIWTGVIGSPA 724
Qy 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
Db 725 ANAPDAGPPQORWFFVW-----LGTANNPVDKGAALAESIRPLVAPPPA----P 770

QY 810 PPPQOGLIPGLMPPSDGSGVTPGTGMPAABMVPPTGSPGGGLPA 856
 DB 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 4

US-10-193-002-209
 ; Sequence 209, Application US/10193002
 ; Publication No. US20030135026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonio
 ; Houghton, Raymond
 ; Vedwick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/193,002
 ; FILING DATE: 10-Jul-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,596
 ; FILING DATE: 05-MAY-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 21021.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 209:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 802 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 209:
 ; US-10-193-002-209

Query Match 53.2%; Score 2700; DB 12; Length 802;
 Best Local Similarity 66.8%; Fred.No. 6.1e-161;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
 QY 1 MGHHHHHHVIDICTSPTWEQAAAEAVQARSDVDIRVARVIEDQMAVDSAGKITVRI 60
 DB 1 MGHHHHHHVIDICTSPTWEQAAAEAVQARSDVDIRVARVIEDQMAVDSAGKITVRI 60
 QY 61 KLEVSFMRPAQRCCKSPSPSPETGAGAGIVATTPASSPTVLAETGSLTLLYPLFNWLG 120
 DB 61 KLEVSFMRPAQRP-R-GSKPSPSPETGAGAGIVATTPASSPTVLAETGSLTLLYPLFNWLG 119
 QY 121 PAFHERYPNVTTAQTGSGAGIAQAAGTAVNIGASDAYLSEGDMAAHKGLNIALAIS 180
 DB 120 PAFHERYPNVTTAQTGSGAGIAQAAGTAVNIGASDAYLSEGDMAAHKGLNIALAIS 179

QY 181 QCVNPNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
 DB 180 QCVNPNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239
 QY 241 GDTFLTQYLSKQDPEGKSPGFGTTVPFPAVPGALGENGNGMVTGCAETPGCVAYTG 300
 DB 240 GDTFLTQYLSKQDPEGKSPGFGTTVPFPAVPGALGENGNGMVTGCAETPGCVAYTG 299
 QY 301 ISFLDQASQRLGEAOLGNSSGNFLPDQOSIQAAAAGFASKTPANQALSMIDGPADGY 360
 DB 300 ISFLDQASQRLGEAOLGNSSGNFLPDQOSIQAAAAGFASKTPANQALSMIDGPADGY 359
 QY 361 PIINYEYAI VNNRQKDAATQTLQAFHLHWAITDGNKASFLDQVHFQPLPFAVVKLS DALI 420
 DB 360 PIINYEYAI VNNRQKDAATQTLQAFHLHWAITDGNKASFLDQVHFQPLPFAVVKLS DALI 419
 QY 421 ATISSAEMKTDAAATLAQEAENFERISGDLKTQIDQVESTAGSLQGWGAGTAAGAAV 480
 DB 420 ATISSAEMKTDAAATLAQEAENFERISGDLKTQIDQVESTAGSLQGWGAGTAAGAAV 479
 QY 481 RFQEAANKQKQELDEISTNIRQAGVQYSRADDEEQQAALSSQMGFTQSTVTVDQBELNR 540
 DB 480 RFQEAANKQKQELDEISTNIRQAGVQYSRADDEEQQAALSSQMGFTQSTVTVDQBELNR 537
 QY 541 ANEVEAPMADPTDPTVITPCELTAAKVAQOLVLSADNMREYLAAGAKERQRLATSLRNA 600
 DB 538 PPAPATPVAPPPPPAAANTPNAQPCDPNAPPPADPNAPPPPPVAPNAPQVPR----- 589
 QY 601 AKAYGEVDEEAATALDNDGEGTQAESAGAVGDS-----SABLTDTPRVATAGBNF- 653
 DB 590 -----IDNPVGGSFALPAGWVESDAAHFDYGSALLS-----KTTGDPPPP 630
 QY 654 -----MDLXEAARKLETGQOGASLAHFAFGDGMNTFNLTLQGDYKFRFGPD 697
 DB 631 GQPPPVANDTRIVGLRDLQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
 QY 698 NWECDAAACEASLDQORQWILHMAKLSAAMAKQAVQALHVWARRHEHTYEDIVGLER 757
 DB 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDFSPKPNQINTGTGIGSPA 734
 QY 758 LYAENPSGARDQILPVVAYEQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
 DB 725 ANAPDAGPPQRFVFW-----LGTANNPVDKGAALAEISIRPLVAPPPA---P 770
 QY 810 PPPQOGLIPGLMPPSDGSGVTPGTGMPAABMVPPTGSPGGGLPA 856
 DB 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 5

US-10-193-002-346
 ; Sequence 346, Application US/10193002
 ; Publication No. US20030135026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonio
 ; Houghton, Raymond
 ; Vedwick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA

Db 120 PAFHERYPNVTITTAQGTGSGAGTAQAAAGTVNIGASDAYLSEGDMAAHKGLNIALAISA 179
QY 181 QQVYNLPVSEHLKNGKVLAAVYGGTITKTWDDPOIALNPGVNLPGTAVVPLHRSOGS 240
Db 180 QQVYNLPVSEHLKNGKVLAAVYGGTITKTWDDPOIALNPGVNLPGTAVVPLHRSOGS 239
QY 241 GDTFLEFQYLSKODPEGKSGPFGTITVDFPVPVPGALGNGGVMWTCGAETPGCVAYIG 300
Db 240 GDTFLEFQYLSKODPEGKSGPFGTITVDFPVPVPGALGNGGVMWTCGAETPGCVAYIG 299
QY 301 ISFLDQASQGLGEAOLGNSGNFLPLDQASTQAAAAGFASKTPANCAISMTDGPAPGY 360
Db 300 ISFLDQASQGLGEAOLGNSGNFLPLDQASTQAAAAGFASKTPANCAISMTDGPAPGY 359
QY 361 PIINYEYAI VNNRQKDAATQTLQAEFLHWAITDGNKASFLDQVHFQPLPAAVVKLSDALI 420
Db 360 PIINYEYAI VNNRQKDAATQTLQAEFLHWAITDGNKASFLDQVHFQPLPAAVVKLSDALI 419
QY 421 ATISSAEMKTDATLAQAGNFERISGDLKTQIDQVESTAGSLOGWRGAAGTAAQAAV 480
Db 420 ATISSAEMKTDATLAQAGNFERISGDLKTQIDQVESTAGSLOGWRGAAGTAAQAAV 479
QY 481 RFQEAANKQKQELDEISTNIRQAGVOYSRADEEQOQALSSQMGFV--PTTAASPPSTAAA 537
Db 480 RFQEAANKQKQELDEISTNIRQAGVOYSRADEEQOQALSSQMGFV--PTTAASPPSTAAA 537
QY 541 ANEVEAPMADPPTDVITPCBELTAANKAAQOLVLSADNMREYLAAGAKERQRLATSLRNA 600
Db 538 PPAPATPVVAPPPAAAANTNAQPGDNAAAPPADPNAPPPVIAPNAPQFVR----- 589
QY 601 AKAYGEVDEEAATALDNDGEGTVQAESAGAVGDS-----SAELTDTRVATAGEPNF- 653
Db 590 -----INPVGGSFALPAGWBSDAHFYGSALLS-----KITGDPFPF 630
QY 654 -----MDLKEARKLETGQOASLAHFAFGWNTFNLTLOGDVKRFRGPNWEGDAA 697
Db 631 GQPPPVANDTRIVGLRDLQKLYASAEATDSKAAA-----RLGSDMGFEF--YM 675
QY 698 NWEGDAATACEASLDQORQWILHMAKLSAAMAKQAQVVAQLHWARREHPTIEDIVGLER 757
Db 676 PYGTGTRINQETVSLD-----ANGVSGSASYEYVFRFSDPSKPNQIWTGVIGSPA 724
QY 758 LYAENPSARDQILPVVAYEQORSEKVLTEYNN-----KAALFVNP-PKPPPAIKIDP 809
Db 725 ANAPDAGPPQWFEVW-----LGTANNFVDKGAALAEISIRPLVAPPPA---P 770
QY 810 PPPQOGLIPGLFMPSPDSGVTGCTGMPAAPVMPPTGSPGGGLPA 856
Db 771 APAPASFA-----PAPAPAGEVAP-----TPTTPTPORTLPA 802

RESULT 7

US-10-084-843-184

; Sequence 184, Application US/10084843

; Publication No. US20030143243A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.

; Dillon, David C.

; Campos-Neto, Antonio

; Houghton, Raymond

; Vedvick, Thomas S.

; Twardzik, Daniel R.

; Lodes, Michael J.

; Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

; AND DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 355

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; City: Seattle

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 184:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 184:

US-10-084-843-184

Query Match

Best Local Similarity 46.8%; Score 2375; DB 12; Length 460;

Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 525 TQSTVTVDQOEILNRANEVEAPMADPPTDVITPCBELTAANKAAQOLVLSADNMREYLA 584
Db 2 TQSTVTVDQOEILNRANEVEAPMADPPTDVITPCBELTAANKAAQOLVLSADNMREYLA 61
QY 585 AGAKERQRLATSLRNAKAYGEVDEEAATALDNDGEGTVQAESAGAVGDS SAEITDTPR 644
Db 62 AGAKERQRLATSLRNAKAYGEVDEEAATALDNDGEGTVQAESAGAVGDS SAEITDTPR 121
QY 645 VATAGEPNFMDLKEARKLETGQOASLAHFAFGWNTFNLTLOGDVKRFRGPNWEGDAA 704
Db 122 VATAGEPNFMDLKEARKLETGQOASLAHFAFGWNTFNLTLOGDVKRFRGPNWEGDAA 181
QY 705 TACEASLDQORQWILHMAKLSAAMAKQAQVVAQLHWARREHPTIEDIVGLERLYAENPS 764
Db 182 TACEASLDQORQWILHMAKLSAAMAKQAQVVAQLHWARREHPTIEDIVGLERLYAENPS 241
QY 765 ARDQILPVVAYEQORSEKVLTEYNNKAALFVNPVPPPAIKIDPPPPPOEQGLIPGFLM 824
Db 242 ARDQILPVVAYEQORSEKVLTEYNNKAALFVNPVPPPAIKIDPPPPPOEQGLIPGFLM 301
QY 825 PPSDGSVPTGCTGMPAAPVMPPTGSPGGGLPADTAAQLTSAGREAAALSGDVAVKAAASLG 884
Db 302 PPSDGSVPTGCTGMPAAPVMPPTGSPGGGLPADTAAQLTSAGREAAALSGDVAVKAAASLG 361
QY 885 GGGGGGVSPAPLGSAGSAGSVRPAGAGDIAGLGGQAGGAGALGGGGMGMPGAAHQOQ 944
Db 362 GGGGGGVSPAPLGSAGSAGSVRPAGAGDIAGLGGQAGGAGALGGGGMGMPGAAHQOQ 421
QY 945 GGAKSKGSQDEALYTEDRAWTEAVIGNRRRRQDSKESK 983
Db 422 GGAKSKGSQDEALYTEDRAWTEAVIGNRRRRQDSKESK 460

RESULT 8

US-10-193-002-179

; Sequence 179, Application US/10193002

; Publication No. US20030135026A1

; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 179:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 179:
US-10-193-002-179
Query Match 46.8%; Score 2375; DB 12; Length 460;
Best Local Similarity 100.0%; Pred. No. 6.9e-141;
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 525 TQSQTVTVDQBIILNRAVEAPMADPTDPTIPCELTAANKAAQQLVLSADNNREYLA 584
DB 2 TQSQTVTVDQBIILNRAVEAPMADPTDPTIPCELTAANKAAQQLVLSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAKAYGVDEEATAALDNDGEGTVQAESAGAVGDSASLLTDPTR 644
DB 62 AGAKERQRLATSLRNAKAYGVDEEATAALDNDGEGTVQAESAGAVGDSASLLTDPTR 121
QY 645 VATAGEPNFMDLKEARKLETGDCASLAHFDAGWNNTNLTQGDVKEFRGFDNWDGAA 704
DB 122 VATAGEPNFMDLKEARKLETGDCASLAHFDAGWNNTNLTQGDVKEFRGFDNWDGAA 181
QY 705 TACEASLDQOROWILHMAKLSAAMAKQCYVAQLHWARREHTPTVEDIVGLERLYAENPS 764
DB 182 TACEASLDQOROWILHMAKLSAAMAKQCYVAQLHWARREHTPTVEDIVGLERLYAENPS 241
QY 765 ARDQILPVYAEYQQRSEKVLTEYNNKAALPEVNPPEKPPPAIKIDPPPPPPQEQGLIPGLM 824
DB 242 ARDQILPVYAEYQQRSEKVLTEYNNKAALPEVNPPEKPPPAIKIDPPPPPPQEQGLIPGLM 301
QY 825 PPSDGSVTPGTGMPAAPVPTGSPGGGLPADTAAQLTSAGREAAALSGDVAVKAASLG 884

DB 302 PPSDGSVTPGTGMPAAPVPTGSPGGGLPADTAAQLTSAGREAAALSGDVAVKAASLG 361
QY 885 GGGGGVPSAPLGSATGGAESVRPAGAGDIAGLGGAGGAGALGGGCMGMPGAAHOGQ 944
DB 362 GGGGGVPSAPLGSATGGAESVRPAGAGDIAGLGGAGGAGALGGGCMGMPGAAHOGQ 421
QY 945 GGAKSQSQOQDEALYTEDRAWTEAVIGNRRRQDSKESK 983
DB 422 GGAKSQSQOQDEALYTEDRAWTEAVIGNRRRQDSKESK 460
RESULT 9
US-10-084-843-355
Sequence 355, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 355:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 355:
US-10-084-843-355

Query Match 36.3%; Score 1839; DB 12; Length 652;
Best Local Similarity 48.5%; Pred. No. 4.4e-107;
Matches 441; Conservative 26; Mismatches 132; Indels 310; Gaps 19;
QY 1 MGHHHHHVHDIITGTGTSWEQAABAVQARDSDIRVARVIEQDMAVDSAGKITRYI 60
DB 1 MGHHHHH-----8
QY 61 KLEVSFKMRPAQPRCGSKPPSGFETGAGAGTVATTASSPVTIAETGTLTLLPLNLWG 120

Db 9 -----GSKPSPGSPETGAGATVATTPASSPVTLAETGTLPLPLNLWG 53
QY 121 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNIGASDAYLSGDMAAHKLNNIALAISA 180
Db 54 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNIGASDAYLSGDMAAHKLNNIALAISA 113
QY 181 QQVNYNLPVSEHLKNGKVLAAVYOGTITKTWDDPOIALNPGVNLPGTAVVPLHRS DGS 240
Db 114 QQVNYNLPVSEHLKNGKVLAAVYOGTITKTWDDPOIALNPGVNLPGTAVVPLHRS DGS 173
QY 241 GDTFLFTQYLSKQDPPEGWKS PGFGTTVPFPAVPGALGENGNGMVTCGAETPGCVAYIG 300
Db 174 GDTFLFTQYLSKQDPPEGWKS PGFGTTVPFPAVPGALGENGNGMVTCGAETPGCVAYIG 233
QY 301 ISFLDQASORGLGEAQLGNSSGNFLPDAQSIOAAAAGFASKTPANOALSMIDGPADGY 360
Db 234 ISFLDQASORGLGEAQLGNSSGNFLPDAQSIOAAAAGFASKTPANOALSMIDGPADGY 293
QY 361 PIINYEYAI VNNRQKDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPFAVVKLS DALI 420
Db 294 PIINYEYAI VNNRQKDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPFAVVKLS DALI 353
QY 421 ATISSAEMKTDAA TLQAEGAGNERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAV 480
Db 354 ATISSGG-----GSGGSGSGSGSGSVPTTAASPPSTA-----AAPPAPATPVA 396
QY 481 RFQEAANKQKQELDETSTNIROAGVQYSRADBEQQALSSQMGFTQSTVTVDQCEILNR 540
Db 397 PPPAA-----ANTPNAQPG----- 411
QY 541 ANEVEAPMADPPDVTPIITCELTAAKAAQQLVLSADN-----MREYLAAGAXERQ----- 591
Db 412 -----DENAAPPADPNAPPVPIAENAPQPVRI--DNPVGGFSFALPAGWVESDAAHFD 464
QY 592 -----RLATSLRNAKAYGEVDEEAATALDND-- 618
Db 465 YGSALLSKTGDPPFGQPPPVANDTRIVGLRLDQKLYASAEA---TSSKAALRGLSDMG 521
QY 619 -----GEGTVOAESAGAVGDDSAELTDTFRVATAGEPNFMDLKEAARKLETGD 667
Db 522 EFMYPYPTGRINQETVSLDANGVSGASYEV-----KFSDPKSKNGQIWTGV 569
QY 668 QGASLAHFAQDQWNTNLITLQGDVKRGRFDNHEGDAATACEASLQOQRCWILHMAKLSAA 727
Db 570 IGSPAANAFDA-----GPPQRW--FVVMGLGTANNPVDK----- 600
QY 728 MAKQAOYVAQLHWARREHPTVEDIVGLERLYAENPSARDQILPVYAYQORSEKVLTEY 787
Db 601 -----GAAKALAE--SIRPLVAP----- 616
QY 788 NNKAALBPVNPKPAPPAIKIDPPPPQEOGLIPGFLMPPSDSGSVTPTGMPAPAPMVPPT 847
Db 617 -----PPAPAPA-PAEPAPAPAPAGEV-----AP-TPTT 643
QY 848 GSPGGLPA 856
Db 644 PTFORTLPA 652

RESULT 10

US-10-193-002-350
; Sequence 350, Application US/10193002
; Publication No. US20030135026A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, David C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.

; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 350:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 350:
US-10-193-002-350

Query Match 36.3%; Score 1839; DB 12; Length 652;
Best Local Similarity 48.5%; Pred. No. 4.4e-107;
Matches 441; Conservative 26; Mismatches 132; Indels 310; Gaps 19;
QY 1 MGHHHHHHVIDIGTSTSWEQAAAEAVQARSDVDIRVARVIEQDMAVDSAGKITRYI 60
Db 1 MGHHHHHH----- 8
QY 61 KLEVSFKWRPAQPRCGSKPPSGSPETGAGATVATTPASSPVTLAETGTLPLPLNLWG 120
Db 9 -----GSKPSPGSPETGAGATVATTPASSPVTLAETGTLPLPLNLWG 53
QY 121 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNIGASDAYLSGDMAAHKLNNIALAISA 180
Db 54 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNIGASDAYLSGDMAAHKLNNIALAISA 113
QY 181 QQVNYNLPVSEHLKNGKVLAAVYOGTITKTWDDPOIALNPGVNLPGTAVVPLHRS DGS 240
Db 114 QQVNYNLPVSEHLKNGKVLAAVYOGTITKTWDDPOIALNPGVNLPGTAVVPLHRS DGS 173
QY 241 GDTFLFTQYLSKQDPPEGWKS PGFGTTVPFPAVPGALGENGNGMVTCGAETPGCVAYIG 300
Db 174 GDTFLFTQYLSKQDPPEGWKS PGFGTTVPFPAVPGALGENGNGMVTCGAETPGCVAYIG 233
QY 301 ISFLDQASORGLGEAQLGNSSGNFLPDAQSIOAAAAGFASKTPANOALSMIDGPADGY 360
Db 234 ISFLDQASORGLGEAQLGNSSGNFLPDAQSIOAAAAGFASKTPANOALSMIDGPADGY 293
QY 361 PIINYEYAI VNNRQKDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPFAVVKLS DALI 420
Db 294 PIINYEYAI VNNRQKDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPFAVVKLS DALI 353
QY 421 ATISSAEMKTDAA TLQAEGAGNERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAV 480

Db 2 KIRLHTLLAVLTAAPIILLAAAGCGSPSSPETGAGAGTAVTTTASSPVTLAETGSTLL 61
QY 113 YPLENLWGPAFHRYPNVITTAQGTSGAGIAQAAAGTNNIGASDAYLSEGDMAAHKGLM 172
Db 62 YPLENLWGPAFHRYPNVITTAQGTSGAGIAQAAAGTNNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAOQVNNYLPVSEHLKLNKGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAV 232
Db 122 NIALAISAOQVNNYLPVSEHLKLNKGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAV 181
QY 233 PLHRSDGSGDTFTFTQYLSKQDPGKSPGFGTTVDFFPAVPCALGENGNGMVTGCAET 292
Db 182 PLHRSDGSGDTFTFTQYLSKQDPGKSPGFGTTVDFFPAVPCALGENGNGMVTGCAET 241
QY 293 PGCVAVIGISFLDOASQORGLGEAQLGNSGNNFLPDAQSIQAAAAGFASKTANQAISMI 352
Db 242 PGCVAVIGISFLDOASQORGLGEAQLGNSGNNFLPDAQSIQAAAAGFASKTANQAISMI 301
QY 353 DGPAPDGYPIINYEYAIVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHFQPLPPAV 412
Db 302 DGPAPDGYPIINYEYAIVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
Db 362 VKLSDALIATISS 374

RESULT 13

US-10-084-843-153
; Sequence 153, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 153:
US-10-084-843-153

Query Match 36.1%; Score 1831; DB 12; Length 374;
Best Local Similarity 95.2%; Pred. No. 6.6e-107;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;
QY 59 RIKLEVSFKWRPQP-----RCGSKPPSPGSPETGAGAGTAVTTTASSPVTLAETGSTLL 112
Db 2 KIRLHTLLAVLTAAPIILLAAAGCGSPSSPETGAGAGTAVTTTASSPVTLAETGSTLL 61
QY 113 YPLENLWGPAFHRYPNVITTAQGTSGAGIAQAAAGTNNIGASDAYLSEGDMAAHKGLM 172
Db 62 YPLENLWGPAFHRYPNVITTAQGTSGAGIAQAAAGTNNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAOQVNNYLPVSEHLKLNKGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAV 232
Db 122 NIALAISAOQVNNYLPVSEHLKLNKGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAV 181
QY 233 PLHRSDGSGDTFTFTQYLSKQDPGKSPGFGTTVDFFPAVPCALGENGNGMVTGCAET 292
Db 182 PLHRSDGSGDTFTFTQYLSKQDPGKSPGFGTTVDFFPAVPCALGENGNGMVTGCAET 241
QY 293 PGCVAVIGISFLDOASQORGLGEAQLGNSGNNFLPDAQSIQAAAAGFASKTANQAISMI 352
Db 242 PGCVAVIGISFLDOASQORGLGEAQLGNSGNNFLPDAQSIQAAAAGFASKTANQAISMI 301
QY 353 DGPAPDGYPIINYEYAIVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHFQPLPPAV 412
Db 302 DGPAPDGYPIINYEYAIVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
Db 362 VKLSDALIATISS 374

RESULT 14

US-10-084-843-155
; Sequence 155, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 155:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 155:
US-10-084-843-155

Query Match 36.1%; Score 1831; DB 12; Length 374;
Best Local Similarity 95.2%; Pred. No. 6.6e-107;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKMPAPQ-----RCGSKPPSGSPETGAGAGTVAITPASSPVTLAETGSTALL 112
DB 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGAGTVAITPASSPVTLAETGSTALL 61
QY 113 YPLFNLMGPAPHERYPNVITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLFNLMGPAPHERYPNVITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAOQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAOQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDGSGDTFLFTQYLSKQDPEGWKGSPGFGTTVDPPVPGALGENGGWMTGCAET 292
DB 182 PLHRSDGSGDTFLFTQYLSKQDPEGWKGSPGFGTTVDPPVPGALGENGGWMTGCAET 241
QY 293 PGCVAYIGISFLDQASQRLGEAQLGNSGNNFLPDQAQSIQAAAAGFASKTPANQAISMI 352
DB 242 PGCVAYIGISFLDQASQRLGEAQLGNSGNNFLPDQAQSIQAAAAGFASKTPANQAISMI 301
QY 353 DGPAPDGYPIINYEYAVNNRQKDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAVNNRQKDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

RESULT 15
US-10-193-002-148
Sequence 148, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:
APPLICANT: Skeedy, Steven G.
Skeedy, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/193,002
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-10-193-002-148

Query Match 36.1%; Score 1831; DB 12; Length 374;
Best Local Similarity 95.2%; Pred. No. 6.6e-107;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKMPAPQ-----RCGSKPPSGSPETGAGAGTVAITPASSPVTLAETGSTALL 112
DB 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGAGTVAITPASSPVTLAETGSTALL 61
QY 113 YPLFNLMGPAPHERYPNVITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLFNLMGPAPHERYPNVITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAOQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAOQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDGSGDTFLFTQYLSKQDPEGWKGSPGFGTTVDPPVPGALGENGGWMTGCAET 292
DB 182 PLHRSDGSGDTFLFTQYLSKQDPEGWKGSPGFGTTVDPPVPGALGENGGWMTGCAET 241
QY 293 PGCVAYIGISFLDQASQRLGEAQLGNSGNNFLPDQAQSIQAAAAGFASKTPANQAISMI 352
DB 242 PGCVAYIGISFLDQASQRLGEAQLGNSGNNFLPDQAQSIQAAAAGFASKTPANQAISMI 301
QY 353 DGPAPDGYPIINYEYAVNNRQKDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAVNNRQKDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

Search completed: November 21, 2003, 16:38:18
Job time : 42.3392 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 20.507 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-54
Perfect score: 5072
Sequence: 1 MGHHHHHVIDIIGTPTSW.....RAWTEAVIGNRRQDSKESK 983

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*1: pir1:*2: pir2:*3: pir3:*4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2375	46.8	460	2 G70803	hypothetical prote
2	1831	36.1	374	1 F70584	phosphate specific
3	485	9.6	100	2 H70802	hypothetical prote
4	482.5	9.5	392	2 AG2377	phosphate-binding
5	463	9.1	364	2 G82595	ABC transporter ph
6	460.5	9.1	373	2 H72756	probable phosphate
7	456	9.0	347	2 AD1920	ABC phosphate tran
8	445	8.8	403	2 H75611	phosphate ABC tran
9	444	8.8	383	1 S74423	phosphate-binding
10	440	8.7	346	2 AI0499	probable phosphate
11	419	8.3	346	2 H91211	phosphate-binding
12	419	8.3	346	2 H85057	phosphate-binding
13	413	8.1	346	1 BYSCPR	phosphate-repressi
14	411	8.1	346	2 AB9556	periplasmic phosph
15	406	8.0	383	2 D75081	phosphate abc tran
16	404	8.0	389	1 F64426	phosphate-binding
17	400	7.9	337	1 C70473	phosphate-binding
18	380.5	7.5	405	2 C90194	hypothetical prote
19	378	7.5	429	2 C87183	phosphate-binding
20	335.5	6.6	369	2 B81171	PstS component of
21	330.5	6.5	370	2 D70584	phosphate uptake s
22	325	6.4	71	2 F70833	probable protein t
23	316.5	6.2	333	1 S74876	phosphate-binding
24	314	6.2	370	2 H70583	phosphate-binding
25	304	6.0	258	1 I64120	phosphate-binding
26	226.5	4.5	326	1 S39852	phosphate-binding
27	220	4.3	298	2 D97110	periplasmic phosph
28	203	4.0	1804	2 H96597	hypothetical prote
29	202.5	4.0	298	2 A86840	hypothetical prote

ALIGNMENTS

RESULT 1

G70803

hypothetical protein Rv3881c - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: G70803

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: G70803

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-460 <COL>

A/Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL1973.1; PID:el26413;

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: Rv3881c

Query Match 46.8%; Score 2375; DB 2; Length 460;

Best Local Similarity 100.0%; Pred. No. 2.9e-106;

Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 525 TQSTVTVDQOEILNRANEVEAPWADPPTDVPITPCELTAAKNAQAOLVLSADNMREYLA 584

Db 2 TQSTVTVDQOEILNRANEVEAPWADPPTDVPITPCELTAAKNAQAOLVLSADNMREYLA 61

Qy 585 AGAKERORLATSRLNAAKAYGEVDEDEAATALDNDGEGTVOAESAGAVGGSSAELDTTPR 644

Db 62 AGAKERORLATSRLNAAKAYGEVDEDEAATALDNDGEGTVOAESAGAVGGSSAELDTTPR 121

Qy 645 VATAGEFNFDLKEARKLETGQGSALAHFADGWNFTNLTDGDKVRPFGFNWEGDAA 704

Db 122 VATAGEFNFDLKEARKLETGQGSALAHFADGWNFTNLTDGDKVRPFGFNWEGDAA 181

Qy 705 TACEASLDQQRQWILHMAKLSAAMAKQAQYVAQLHVWARREHPTYEDIVGLERLYAENPS 764

Db 182 TACEASLDQQRQWILHMAKLSAAMAKQAQYVAQLHVWARREHPTYEDIVGLERLYAENPS 241

Qy 765 ARDQILPVYAEYQORSEKVLTEYNNKALEFVNPPKPPPAIKIDPPPPQOGLIPGFLM 824

Db 242 ARDQILPVYAEYQORSEKVLTEYNNKALEFVNPPKPPPAIKIDPPPPQOGLIPGFLM 301

Qy 825 PPSDGSVTPGTGMPAAPMPVPTGSPGGGLPADTAALTSAGREAAALSGDVAKVASLG 884

Db 302 PPSDGSVTPGTGMPAAPMPVPTGSPGGGLPADTAALTSAGREAAALSGDVAKVASLG 361

Qy 885 GGGCGGVPSPLGSAIGAESVTPAGAGDTAGLQGRAGGAALGGCGMPPGAAHQGQ 944

Db 362 GGGGGVPSAPLGAISGAESVRPAGAGDIAGLQGRAGGAALGGGGMPMGAHQOQ 421

Qy 945 GGAKSQSGQDEALYTEDRAWTEAVIGNRRQDSKESK 983

Db 422 GGAKSQSGQDEALYTEDRAWTEAVIGNRRQDSKESK 460

RESULT 2

F70584

Phosphate specific transporter S precursor - Mycobacterium tuberculosis (strain H37RV)

N;Alternate names: antigen b

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C;Species: Mycobacterium tuberculosis

C;Accession: F70584; JCS103; A42930; A49721; A45820

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70584

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-374 <COL>

A;Cross-references: GB:295209; GB:AL123456; NID:G3261750; PIDN:CAB08484.1; PID:G2078049

A;Experimental source: strain H37RV

R;Braibant, M.; Lefevre, P.; de Wit, L.; Peirs, P.; Ooms, J.; Huygen, K.; Andersen, A.B.

Gene 176, 171-176, 1996

A;Title: A Mycobacterium tuberculosis gene cluster encoding proteins of a phosphate transporter

A;Reference number: JCS100; MUID:97075926; PMID:8918249

A;Accession: JCS103

A;Molecule type: DNA

A;Residues: 1-374 <BRA>

A;Cross-references: GB:M30046; NID:G149987; PIDN:AAA25374.1; PID:G149988

A;Note: neither the complete nucleic acid sequence nor the complete translation are shown

R;Andersen, A.B.; Hansen, E.B.

Infect. Immun. 57, 2481-2488, 1989

A;Title: Structure and mapping of antigenic domains of protein antigen b, a 38,000-molecular weight protein

A;Reference number: A42930; MUID:89307568; PMID:2545626

A;Accession: A42930

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-374 <AND1>

A;Cross-references: GB:M30046; NID:G149987; PIDN:AAA25374.1; PID:G149988

R;Chang, Z.; Choudhary, A.; Lathigra, R.; Quijcho, F.A.

J. Biol. Chem. 269, 1956-1958, 1994

A;Title: The immunodominant 38-kDa lipoprotein antigen of Mycobacterium tuberculosis is

A;Reference number: A49721; MUID:94124544; PMID:8294447

A;Accession: A49721

A;Molecule type: protein

A;Residues: 25-34 <CHA>

R;Andersen, A.B.; Ljungqvist, L.; Olsen, M.

J. Gen. Microbiol. 136, 477-480, 1990

A;Title: Evidence that protein antigen b of Mycobacterium tuberculosis is involved in phagocytosis

A;Reference number: A45820; MUID:90362031; PMID:2118164

A;Accession: A45820

A;Molecule type: protein

A;Residues: X', 27-28, X', 30-34, XX', 37 <AND2>

A;Note: confirmed presence of normal signal and absence of lipoprotein attachment

C;Genetics:

A;Gene: phoS1; pats

A;Start codon: GTG

C;Superfamily: phosphate-repressible phosphate-binding protein

C;Keywords: surface antigen

F;1-25/Domain: signal sequence #status predicted <SIG>

F;26-374/Product: phosphate specific transporter S #status experimental <MAT>

Query Match 36.1%; Score 1831; DB 1; Length 374;

Best Local Similarity 95.2%; Pred. No. 1.9e-80;

Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

Qy 59 RIKLEVSFKMRPAOP-----RCGSKPPSGSPETGAGAGTVAATPASSPVTLAETGSTALL 112

Db 2 KIRLHTLLAVITAAPLLAAAGCGSKPPSGSPETGAGAGTVAATPASSPVTLAETGSTALL 61

Qy 113 YPLFNLWGPAFHERYPNVTTTAQCTGSGAGIAQAAGATVNVIGASDAYLSEGDVAAHKGML 172

Db 62 YPLFNLWGPAFHERYPNVTTTAQCTGSGAGIAQAAGATVNVIGASDAYLSEGDVAAHKGML 121

Qy 173 NIALAISAOQVNNLPGVSEHLKNGKVLAAVGGTITKTDWDDPOIALNPGVNLPGTAVV 232

Db 122 NIALAISAOQVNNLPGVSEHLKNGKVLAAVGGTITKTDWDDPOIALNPGVNLPGTAVV 181

Qy 233 FLHRSDSGDTFLFTQYLSKQDPGKSGFPGTITVDFPAVPGALGNGGMMVTGCAET 292

Db 182 FLHRSDSGDTFLFTQYLSKQDPGKSGFPGTITVDFPAVPGALGNGGMMVTGCAET 241

Qy 293 PGCVAYTIGISFLDOASORGLEAGLGNSSGNFLPDAQSTQAAAAGFASKTPANQAISMI 352

Db 242 PGCVAYTIGISFLDOASORGLEAGLGNSSGNFLPDAQSTQAAAAGFASKTPANQAISMI 301

Qy 353 DGPAPDGYPIINYEYATVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHQPPLPAV 412

Db 302 DGPAPDGYPIINYEYATVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHQPPLPAV 361

Qy 413 VKLSDALIATISS 425

Db 362 VKLSDALIATISS 374

RESULT 3

H70802

Hypochemical protein Rv3874 - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: H70802

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-100 <COL>

A;Cross-references: GB:AL022120; GB:AL123456; NID:G3261558; PIDN:CAA17966.1; PID:G2960222

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: Rv3874

Query Match 9.6%; Score 485; DB 2; Length 100;

Best Local Similarity 100.0%; Pred. No. 5.5e-17;

Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 426 AEMKTDATLAQAGNPERISGDLKTDQVSTAGSLQGWGAGTAACAQAAVVRFOEA 485

Db 2 AEMKTDATLAQAGNPERISGDLKTDQVSTAGSLQGWGAGTAACAQAAVVRFOEA 61

Qy 486 ANKOKQLDEISTNIRQAGVQYGRADDEEQQAALSSQMGF 524

Db 62 ANKOKQLDEISTNIRQAGVQYGRADDEEQQAALSSQMGF 100

RESULT 4

AG2377

Phosphate-binding periplasmic protein of phosphate ABC transporter all4575 [imported] - N

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AG2377

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2377
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-392 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA876274.1; PID:gl7133711; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all4575
C;Superfamily: phosphate-repressible phosphate-binding protein

Query Match 9.5%; Score 482.5; DB 2; Length 392;
Best Local Similarity 33.7%; Pred. No. 4.1e-16;
Matches 124; Conservative 64; Mismatches 151; Indels 29; Gaps 12;

QY 75 CGSKPPSPGSPETGAG-AGTAVATTPASSP-----VTLAETGTTLLYPLNMGPAFHE 125
DB 29 CGGQGSNTATQDSSGTAXDATASSPAKLDLGNVSLTGAGASFPAPLYASWFTLNK 88

QY 126 RYPNVTTTAQGTGSGAGTAAAGTAVNIGASDAYLSEGDMAHKLNMNIALAISAOQVNV 184
DB 89 KYPNLQINQVSGGAGVEQTIQGTVDVFGASDVAMKDEBIQVQGVLLFPVTTAGGIVLA 148

QY 185 YNLPGVSEHLKINGKVLAAVYQGTIKTWDWDPQIAALNPGVNLPGTAVVPLHRS DSGSDT 244
DB 149 YNLPQVTE-LNLPRVYTDILLGLIKTWDAPBEIKAAFNPNLPSQPTIVVRS DSGSTTG 207

QY 245 LFTQVLSKODPEGWKSFGFTTVDPAVPGALGNGNGMVTGCATPFGCAVIGISFL 304
DB 208 VFTKHLAAVSP-EWKSQVGEKSVSWPV---GVGKGNGEVTAQIKOTQGAIGVIEGY- 262

QY 305 DOASQGLGEAQLGNSCNELLPPAQSIOAAAAGFASKTPANOAISMDGPAPDG---YP 361
DB 263 --AKONNISYATLENKAGKFKYKINDESASQTIA--ALQLENLRAFY---PDPDGDGSYP 315

QY 362 IINYEYAIWNRQKDAATQTLQAFHLHWAITDGNKASFLDQVHFQPLPPAVVKS DALLIA 421
DB 316 IVSFSWIMAYKNYPDPVKAKAMEAIEYALTEGQKIS--GELGYILPQAVVQ-KTATVA 372

QY 422 TISSNEMK 429
DB 373 DQISPEYK 380

RESULT 5
ABC transporter phosphate binding protein XF2141 [imported] - Xylella fastidiosa (strain
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: G82595
R;anonymouse, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: G82515; MUID:20365717; PMID:10910347
A;Note: For a complete list of authors see reference number A59328 below
A;Accession: G82595
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <SIM>
A;Cross-references: GB:AE004028; GB:AE003849; NID:g9107266; PIDN:AAF84940.1; GSPDB:GN001
A;Experimental source: strain 945c
R;Simpton, A.J.O.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A;Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Vertovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XP2141
C;Superfamily: phosphate-repressible phosphate-binding protein

Query Match 9.1%; Score 463; DB 2; Length 364;
Best Local Similarity 32.8%; Pred. No. 3.2e-15;
Matches 116; Conservative 67; Mismatches 151; Indels 20; Gaps 10;

QY 76 GSKPPSPGSPETGAGAGTAVATTPASSP-----PVTLAETGTTLLYPLNMGPAFHERYPNV 131
DB 21 GCKPSNDNOSTGVSDGNSITPPSAEQTAKSVKISGAGASFIPLISQWADINAATGN-K 79

QY 132 ITAQGTGSGAGTAAAGTAVNIGASDAYLSEGDMAHKLNMNIALAISAOQVNVNLP 191
DB 80 INYQISGSGGTAQIAKAAITDFGSSDKPLDSSSEL-TQAGLGQFPSPAIGGVPWVNLNIE 138

QY 192 E-HKLNGKVLAAVYQGTIKTWDWDPQIAALNPGVNLPGTAVVPLHRS DSGSDTFLFTQYL 250
DB 139 PGKLRUTGPELLADIFLGKISKWNDAAIISANPGLHLPTKINIVHRS DSGSTGFNFNSYL 198

QY 251 SKQDPSEGWKSFGFTTVDPAVPGALGNGNGMVTGCATPFGCAVYIGISFLDQASOR 310
DB 199 SKVSAB-WKQVGEGETSVQV---PGVGKGNGEVASYVQOIKSGSIGYVELAV---ALQN 251

QY 311 GLGEAQLGNSCNELLPPAQSIOAAA--AGFASKTPANOAISMDGPADGPIINYEVA 368
DB 252 KMSYTTALNAAGQWQVQPSAESFAAASNADWNAKDFNLVITNATGEA--AWPITATNFI 309

QY 369 IYNRQKDAATQTLQAFHLHWAITDGNKASFLDQVHFQPLPPAVVKS DALLIAT 422
DB 310 LMRKQTKDAQKAKATLDFKNSFENGQKQA--NELHYVLPPLNVLKQIEAYWAS 361

RESULT 6
H72756
Probable phosphate-binding periplasmic protein APE0045 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: H72756
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: H72756
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-373 <RAW>
A;Cross-references: DDBJ:APC00058; NID:g5103388; PIDN:BAA78954.1; PID:g5103433
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0045
C;Superfamily: phosphate-repressible phosphate-binding protein

Query Match 9.1%; Score 460.5; DB 2; Length 373;
Best Local Similarity 32.0%; Pred. No. 4.3e-15;
Matches 112; Conservative 65; Mismatches 150; Indels 23; Gaps 9;

QY 85 ETGAGTAVATTPASSPVTLAETGTTLLYPLNMGPAFHERYPNVTTTAQGTGSGAGIA 144
DB 34 ESSGGSG-----PAQ--VVVQEGSGTFIYPOIQAMSEEEKTPYTPMTTINYNPTGSGAGOS 86

QY 145 QAAAGTAVNIGASDAYLSEGDMAHKG-LMNIALAISAOQVNVNLPVSHLKLNGVLA 203
DB 87 AFQGVVDVFGSDTLPDPREVWISANQGVNIQIPVIGMLAIVNIPGL-EOLKLDATIAL 145

QY 204 MYQGTIKTWDWDPQIAALNPGVNLPGTAVVPLHRS DSGSDTFLFTQVLSKODPEGWKS 263
DB 146 IYKGEIYWDPRIASINLPGASLPHEKIVAVHRS DSGSTHTVFTIFLHKGADPVWTEDL- 204

QY 264 FGTTVDFPP--AVPGALGNGNGMVTGCAETPCVAYIGISFLDQASQRLGAEQAGNS 321
Db 205 VGKSVDFVDPATGKGVGKNGQVMEVWNTPTYSIGVEYAYVVKAGE-GVNVALVNRD 263

QY 322 GNELLDPAQSIQAAAGFASKTP-----ANQAI SMIDGPADGYPPIINYEYAVNNRQ 374
Db 264 DVFLKSPGGAQAASAGAVNLPDSDPDWSTGYDAIIYAPGKDSYPIITWSFLFYKQY 323
QY 375 KDAATQATLQAFHLHWAITDGNKASFLDQVHFQPLPPAVVVKLSDALIATIS 424
Db 324 NDRDKAEAIKKEIWEINTEGQTIIIEGYI---PIPDEIRQINKKAVEMIS 370

RESULT 7
AD1920
ABC phosphate transport system phosphate-binding periplasmic protein all0911 [imported]
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 03-Dec-2002
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1920
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-347 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072868.1; PID:gl7130257; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0911
C:Superfamily: phosphate-repressible phosphate-binding protein

Query Match 9.0%; Score 456; DB 2; Length 347;
Best Local Similarity 33.0%; Pred. No. 6.5e-15;
Matches 114; Conservative 69; Mismatches 142; Indels 20; Gaps 10;

QY 86 TGAGACTVATTPASSPVTLAET-----GSTLLYPLFLNLWGPAPHERYPNVITTAQGTGSGA 141
Db 14 TAVVTSVAVVTPIPTAQAQTLNGAGATPPAPLYERYAREVRKKGPELRINYYAIGSGG 73
QY 142 GIAQAAGATVNICASDAYLSEGDMAAHK-GLMNIALAISAQVNNLPGVSEHLKINGKV 200
Db 74 GIRQTITAGTVDFGSDAAMDADIAIKVKGVLVPTAGGAVSVVYVPGV-NLRLSRAT 132

QY 201 LAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDSGDTFLFTQYLSKQDPEGWKG 260
Db 133 LPAIFSGQITNMNDKIRADNPNGVNLPNOPIRFAVRADSSGTTFIPTNHLSSISPYFKGR 192
QY 261 SPGFGTTVDPPAVPGALGNGNGMVTGCAETPGCVAYIGISFLDQASQRLGAEQAGNS 320
Db 193 -VGANTAPKW-TLPNVLRGNGNPGVVALIARTPGSGYVEYAY--ATQNRLSRQIQNR 247
QY 321 SGNFLPLDAQSIQAAAGFASKTPANQAI SMIDGPADGYPPIINYEYAVNNRQKDAATA 380
Db 248 RGEFVAPSLQSAANAALS--AVSPDNFRFV--GDPGQGYPIVGLTMMVMVRYQYADAAS 303

QY 381 QTLQAFHLHWAITDGNKASFLDQVHFQPLPPAVVVKLSDALIATIS 425
Db 304 QAIKKWNNVLKRGQQ--FNDDLNYYTRIPDAV---ANRVLTQVNS 343

RESULT 8
H75611
phosphate ABC transporter, periplasmic phosphate-binding protein - Deinococcus radiodurans
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lan, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: H75611
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-403 <WHI>
A:Cross-references: GB:AE001862; GB:AE001825; NID:96460469; PIDN:AAF12207.1; PID:96460501
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0157
A:Map position: 2

Query Match 9.8%; Score 445; DB 2; Length 403;
Best Local Similarity 31.5%; Pred. No. 2.6e-14;
Matches 118; Conservative 60; Mismatches 157; Indels 40; Gaps 11;

QY 73 PRGSKPPSGS-----PETGAGACT---VATTPASSPVTLAETGSTLLYP 114
Db 40 PLCGTSPVFGSLDAPLTKPRHTRRMKTLILGLSALVYISTAAQAAGAITGA--GASPPYP 97
QY 115 LFNLMGPAFHERYPNVITTAQGTGAGIAQAAGATVNICASDAYLSEGDMAAHK-IMN 173
Db 98 LYS---KMFSEYKASNVVYQSVGSGSQKILERTVDFAGSDNPMTDAQLGSAFGTLH 153
QY 174 IALAIQAQVNNLPGVSEHLKINGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVP 233
Db 154 VPTAIGAVVPAYNLPGVTKPLNFDGPTLANIYLKIKTWGDPALAKLNPVTPPLPITV 213
QY 234 LHRSDGSGDTFLFTQYLSKQDPEGWKSQPGTGTVDPPAVPGALGNGNGMVTGCAETP 293
Db 214 ARSDGSGTTFVPSDYLKVSQGE-WKSKVAGNSLQWPGVGTGAKNGDGVAGVKG---TP 269

QY 294 GCVAYIGISFLDQASQRLGAEQAGNSGNFLPLDAQSIQAAAGFASKTPANQAI SMID 353
Db 270 GAIGYVELVY---AKQNKLSFGAVKVRAGKFFILADNGFASNAALGVV--IPADTRVSLTN 324
QY 354 GPAPDGPPIINYEYAVNNRQK---DAATAQTLOAFHLHWAITDGNKASFLDQVHFQPLP 409
Db 325 SANAGAVPIASFTYLIIFYKQKYGNRTAQAQKALKNLLTYVTSGQQ--YNEGLDYAKLP 382

QY 410 PAVVVKLSDALIATIS 424
Db 383 SNVAAKAKTIINSWN 397

RESULT 9
S74423
phosphate-binding periplasmic protein pstS-1 - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein sl10680
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 10-Sep-1993 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74423
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-383 <KAN>
A:Cross-references: EMBL:D64001; GB:AB001339; NID:gl001102; PIDN:BA010341.1; PID:gl001197
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
C:Superfamily: phosphate-repressible phosphate-binding protein

Query Match 8.8%; Score 444; DB 1; Length 383;
Best Local Similarity 31.5%; Pred. No. 2.7e-14;

C;Superfamily: phosphate-repressible phosphate-binding protein

Query Match 8.3%; Score 419; DB 2; Length 346;
Best Local Similarity 32.5%; Pred. No. 3.7e-13;
Matches 109; Conservative 56; Mismatches 144; Indels 26; Gaps 10;
QY 92 TVATTPAS-----SPVTLAETGSTALLYPLNMGPAHERYVNVITTAQGTGSGA 141
DQ 7 TVATVVAATLSMSAFSAFASLTGAGATPPAPVYAKWADTYOKETGN-KVNYQGIGSSG 65
QY 142 GIAQAAAGTVNIGASDAYLSEGDVAAHKLMTALALSAQQVNNYLPV-SEHLKNGKV 200
DQ 66 GVKQITANTVDFGASDAPLSD-EKLAQEGLPQPTVIGGVVLAVNIPGLKSGELVDGKT 124
QY 201 LAAMYQGTITKWDPPQIAALNPGVNLPGTAVVPLHRSDDSGDTFLFTQYLSKQDPGSGWK 260
DQ 125 LGDIYLGKIKKWDDEATAKLNPKLPSONIAVVRADSGTSFVFTSYLAKVNEE-WKN 183
QY 261 SPFGTITVDPAVPGALGENGMVTCGAETGCVAYIGISFLDOASQRLGEAQLGNS 320
DQ 184 NVGTGSTVKWPI---GLGKGNDAIAFVORLPGAIGYVEYAY---AKONNLAYTKLISA 237
QY 321 SGNFLLPDAOSIQAAAAGFA-SKTPANQAISMTDGPAPGCPYIINVEYAVVNNRQKDAAT 379
DQ 238 DGKPVLPTEENFANAAGADWSKTFQA---DLTNQKGEDAWPITSTFILIHKKDKKPEQ 294
QY 380 AQTLQAFHLWAITDGNKASFLDQVHFQPLPPAVVK 414
DQ 295 GTEVLKFFDWAYKTGAQKA--NDLDYASLPDSVVE 327

RESULT 13

BVECEP

phosphate-repressible phosphate-binding protein precursor. [validated] - Escherichia coli

C;Species: Escherichia coli
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 01-Mar-2002
C;Accession: A30277; A65176
R;Surin, B.P.; Jans, D.A.; Fimmel, A.L.; Shaw, D.C.; Cox, G.B.; Rosenberg, H.
J. Bacteriol. 157, 772-778, 1984
A;Title: Structural gene for the phosphate-repressible phosphate-binding protein of Esch
A;Reference number: A30277; MUID:84135579; PMID:6321434
A;Accession: A30277
A;Molecule type: DNA
A;Residues: 1-346 <SUR>
A;Cross-references: GB:K01992; NID:G147255; PIDN:AAA24378.1; PID:G147256
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A65176
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-346 <BLAT>
A;Cross-references: GB:AE000449; GB:U00096; NID:G2367269; PIDN:AACT6751.1; PID:G2367271;
A;Experimental source: strain K-12, substrain MG1655
R;Wang, Z.; Luecke, H.; Quioco, F.A.
Submitted to the Brookhaven Protein Data Bank, August 1996
A;Reference number: A68487; PDB:1IXH
A;Contents: annotation; X-ray crystallography, 0.98 angstroms, residues 26-346
R;Yao, N.; Choudhary, A.; Ledvina, P.S.; Quioco, F.A.
Submitted to the Brookhaven Protein Data Bank, November 1995
A;Reference number: A67875; PDB:2ABH
A;Contents: annotation; X-ray crystallography, 1.70 angstroms, residues 26-346
R;Luecke, H.; Quioco, F.A.
Submitted to the Brookhaven Protein Data Bank, April 1992
A;Reference number: A51014; PDB:1ABH
A;Contents: annotation; X-ray crystallography, 1.7 angstroms, residues 26-110, 'G', 112-20
R;Luecke, H.; Quioco, F.A.
Nature 347, 402-406, 1990
A;Title: High specificity of a phosphate transport protein determined by hydrogen bonds.
A;Reference number: A30645; MUID:91015319; PMID:2215649
A;Contents: annotation; X-ray crystallography, 1.7 angstroms

C;Comment: This protein is a component of the inducible, high-affinity, phosphate-specific
C;Genetics:
A;Gene: pstS; phoS
A;Map position: 84 min
C;Superfamily: phosphate-repressible phosphate-binding protein
C;Keywords: phosphate transport
F;1-35/Domain: signal sequence #status predicted <SIG>
F;26-346/Product: phosphate-repressible phosphate-binding protein #status experimental <N
F;35-63.81.160.164,166/Binding site: phosphate (Thr, Ser, Asp, Arg, Ser, Thr) #status ex

Query Match 8.1%; Score 413; DB 1; Length 346;

Best Local Similarity 32.2%; Pred. No. 7.2e-13;
Matches 108; Conservative 56; Mismatches 145; Indels 26; Gaps 10;

QY 92 TVATTPAS-----SPVTLAETGSTALLYPLNMGPAHERYVNVITTAQGTGSGA 141
DQ 7 TVATVVAATLSMSAFSAFASLTGAGATPPAPVYAKWADTYOKETGN-KVNYQGIGSSG 65
QY 142 GIAQAAAGTVNIGASDAYLSEGDVAAHKLMTALALSAQQVNNYLPV-SEHLKNGKV 200
DQ 66 GVKQITANTVDFGASDAPLSD-EKLAQEGLPQPTVIGGVVLAVNIPGLKSGELVDGKT 124
QY 201 LAAMYQGTITKWDPPQIAALNPGVNLPGTAVVPLHRSDDSGDTFLFTQYLSKQDPGSGWK 260
DQ 125 LGDIYLGKIKKWDDEATAKLNPKLPSONIAVVRADSGTSFVFTSYLAKVNEE-WKN 183
QY 261 SPFGTITVDPAVPGALGENGMVTCGAETGCVAYIGISFLDOASQRLGEAQLGNS 320
DQ 184 NVGTGSTVKWPI---GLGKGNDAIAFVORLPGAIGYVEYAY---AKONNLAYTKLISA 237
QY 321 SGNFLLPDAOSIQAAAAGFA-SKTPANQAISMTDGPAPGCPYIINVEYAVVNNRQKDAAT 379
DQ 238 DGKPVLPTEENFANAAGADWSKTFQA---DLTNQKGEDAWPITSTFILIHKKDKKPEQ 294
QY 380 AQTLQAFHLWAITDGNKASFLDQVHFQPLPPAVVK 414
DQ 295 GTEVLKFFDWAYKTGAQKA--NDLDYASLPDSVVE 327

RESULT 14

AB0956

periplasmic phosphate-binding protein [imported] - Salmonella enterica subsp. enterica se
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AB0956
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serova
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AB0956
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD03142.1; PID:G16504777; GSPDB:GN00176
C;Genetics:
A;Gene: STV3925
C;Superfamily: phosphate-repressible phosphate-binding protein

Query Match 8.1%; Score 411; DB 2; Length 346;

Best Local Similarity 32.6%; Pred. No. 9e-13;
Matches 109; Conservative 52; Mismatches 147; Indels 26; Gaps 10;

QY 92 TVATTPAS-----SPVTLAETGSTALLYPLNMGPAHERYVNVITTAQGTGSGA 141
DQ 7 TVATVVAATLSMSAFSAFASLTGAGATPPAPVYAKWADTYOKETGN-KVNYQGIGSSG 65
QY 142 GIAQAAAGTVNIGASDAYLSEGDVAAHKLMTALALSAQQVNNYLPV-SEHLKNGKV 200
DQ 66 GVKQITANTVDFGASDAPLSD-EKLAQEGLPQPTVIGGVVLAVNIPGLKSGELVDGKT 124

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 11.5909 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-54

Perfect score: 5072

Sequence: 1 MGHHHHHHVIDIIGTSPTSM.....RAWTEAVIGNRRQDSKESK 983

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2375	46.8	460	1 MB48 MYCTU	Q933k8 mycobacteri
2	1831	36.1	374	1 PST1 MYCTU	P15712 mycobacteri
3	485	9.6	99	1 CF10 MYCTU	O69739 mycobacteri
4	463	9.1	364	1 PST5 XYLEA	O95bk3 xyliella fas
5	444	8.9	351	1 PST5 RHILQ	Q98f12 rhizobium l
6	413	8.1	346	1 PST5 ECOLI	Q98f12 rhizobium l
7	411.5	8.1	344	1 PST5 PASMU	Q9cnj4 pasteurella
8	384	7.6	334	1 PST5 HAEIN	P45192 haemophilus
9	339.5	6.7	369	1 PST3 MYCAV	O9kk89 mycobacteri
10	335.5	6.6	369	1 PST3 MYCLE	O9cbe5 mycobacteri
11	330.5	6.5	370	1 PST2 MYCTU	O05870 mycobacteri
12	314	6.2	370	1 PST3 MYCTU	O05870 mycobacteri
13	203.5	4.0	1783	1 RAA3 CHLRE	Q86343 mycobacteri
14	189	3.7	1790	1 SEPA EMENI	P78621 emericella
15	184	3.6	99	1 CF10 MYCLE	O33084 mycobacteri
16	184	3.6	300	1 YOGG BACSU	P46338 bacillus su
17	183	3.6	2038	1 FSH DSOME	P13709 drosophila
18	178.5	3.5	853	1 VMTL LAMB	P03736 bacterioph
19	176	3.5	3164	1 TEGU ESV11	P10220 herpes simp
20	175.5	3.5	1733	1 VNVA PRVKA	P33485 pseudorab
21	174	3.4	1238	1 SBCC RHOC	O68032 rhodocact
22	174	3.4	2517	1 NCR2 HUMAN	O9y618 h nuclear r
23	172	3.4	924	1 IF2 MYCLE	O92519 mycobacteri
24	171.5	3.4	368	1 PPEL PSEAB	P35482 pseudomonas
25	171.5	3.4	643	1 TEP4 EPTST	Q90501 eptatretus
26	170	3.4	378	1 CSP PLACB	P08672 plasmodium
27	169	3.3	3178	1 YS99 CAEEL	Q09624 caenorhabdi
28	167.5	3.3	2703	1 NOTC DROME	P07207 drosophila
29	166.5	3.3	1508	1 BCSC XANAC	P58938 xanthomonas
30	163.5	3.2	909	1 SBCC DEIRA	O9rt44 deinoxococ
31	162.5	3.2	1527	1 CAIH MOUSE	P39061 mus musculu
32	162	3.2	2774	1 MAPA RAT	P34926 rattus norv
33	161	3.2	2132	1 PGCA MOUSE	Q61282 mus musculu

34 160 3.2 2090 1 HFC1 MESAU
35 159.5 3.1 864 1 ELS RAT
36 159 3.1 2441 1 CBP MOUSE
37 158.5 3.1 2415 1 PGCA HUMAN
38 157.5 3.1 401 1 CSP PLACG
39 157.5 3.1 1781 1 AK12 HUMAN
40 157 3.1 900 1 IF2 MYCTU
41 157 3.1 1012 1 PHCI MOUSE
42 156.5 3.1 860 1 ELS MOUSE
43 155.5 3.1 518 1 TFM4 DROME
44 155 3.1 747 1 ELS BOVIN
45 155 3.1 1005 1 Y456 CHLTR

ALIGNMENTS

RESULT 1
MB48 MYCTU
ID MB48 MYCTU STANDARD; PRT; 460 AA.
AC Q933k8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antigen MTB48.
GN MTB48 OR RV3881C OR MT3996 OR MTV027.16C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Erdbmann;
RX MEDLINE=21320746; PubMed=11427558;
RA Lodes M.J., Dillon D.C., Mohamath R., Day C.H., Benson D.R.,
RA Reynolds L.D., McNeill P., Sampaio D.P., Skeiky Y.A.W., Badaro R.,
RA Persing D.H., Reed S.G., Houghton R.L.;
RA "Serological expression cloning and immunological evaluation of MTB48,
RA a novel Mycobacterium tuberculosis antigen.";
RA J. Clin. Microbiol. 39:2485-2493(2001).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=9825987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry III C.E., Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickley E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. A processed form is shed into
CC the medium during culture.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

P51611 mesocricetu
Q99372 rattus norv
P45481 mus musculu
P16112 homo sapien
P08674 plasmodium
Q02952 homo sapien
P71613 mycobacteri
Q64028 mus musculu
P54320 mus musculu
P49455 drosophila
P04985 bos taurus
O84462 chlamydia t

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AY029285; AAK31576.1; -
 DR EMBL; AL022120; CAA17973.1; -
 DR EMBL; AE007191; AAK48364.1; -
 DR PIR; G70803; G70803.
 DR TIGR; MT3996; -

DR Tuberculin; Rv3881c; -
 KW Antigen; Complete proteome.
 FT DOMAIN 306 428 GLY-RICH.
 SQ SEQUENCE 460 AA; 47593 MW; 7528743226AD7A71 CRC64;

Query Match 46.8%; Score 2375; DB 1; Length 460;
 Best Local Similarity 100.0%; Pred. No. 1.1e-98;
 Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 525 TQQTVDQOEILNRANEVEAPMADPTDPTPCELTAAKNAQAOLVLSADNNREYLA 584
 Db EMBL; AE007191; AAK48364.1; -

Qy 585 AGAKERORLATSARNAKAYGEVDEEAATLNDGEGTVOAESAGAVGDSGSAELTDTPR 644
 Db 62 AGAKERORLATSARNAKAYGEVDEEAATLNDGEGTVOAESAGAVGDSGSAELTDTPR 121

Qy 645 VATAGEPNFMDLKEARKLTGQGGASLAHFADGWNFTNLTLQGDYKFRFGDNWEGDAA 704
 Db 122 VATAGEPNFMDLKEARKLTGQGGASLAHFADGWNFTNLTLQGDYKFRFGDNWEGDAA 181

Qy 705 TACEASLDQORQWLHMAKLSAMAKQAQYVAQLHWARREHTYEDIVGLRELVAENPS 764
 Db 182 TACEASLDQORQWLHMAKLSAMAKQAQYVAQLHWARREHTYEDIVGLRELVAENPS 241

Qy 765 ARQILPVYAEYQORSEKVLTEYNNKAALPEVNPFPKPKIDPPPPQEQGLIPGFLM 824
 Db 242 ARQILPVYAEYQORSEKVLTEYNNKAALPEVNPFPKPKIDPPPPQEQGLIPGFLM 301

Qy 825 PPSDGSVTPGTGMPAPVPPPTGSGGGLPAPTAQAQLTSAGREAAALSGDVAVKAAASIG 884
 Db 302 PPSDGSVTPGTGMPAPVPPPTGSGGGLPAPTAQAQLTSAGREAAALSGDVAVKAAASIG 361

Qy 885 GGGGGVSPAPLSAIGGAESVAPAGAGDIAGLQGRAGGAGGALGGGGMGMPGAAHQOQ 944
 Db 362 GGGGGVSPAPLSAIGGAESVAPAGAGDIAGLQGRAGGAGGALGGGGMGMPGAAHQOQ 421

Qy 945 GGAKSGSQOEDALYTEDRAWTEAVIGNRRRDSKESK 983
 Db 422 GGAKSGSQOEDALYTEDRAWTEAVIGNRRRDSKESK 460

RESULT 2
 PSTI_MYCTU STANDARD; PST; 374 AA.
 AC P15712; O05868;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphate-binding protein 1 precursor (PBP-1) (Psts-1) (Protein
 DE antigen B) (PAB) (Antigen Ag78).
 GN PSTSI OR PHOSI OR RV0934 OR MT0961 OR MTCV0809.05C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89307568; PubMed=2545626;
 RA Andersen A.B.; Hansen E.B.;
 RT "Structure and mapping of antigenic domains of protein antigen b, a
 RT 38,000-molecular-weight protein of Mycobacterium tuberculosis.";
 RL Infect. Immun. 57:2481-2488(1989).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 EX MEDLINE=96295987; PubMed=9634230;
 RA Cole S.T.; Brosch R.; Parkhill J.; Garnier T.; Churcher C.; Harris D.,
 RA Gordon S.V.; Eigemeier K.; Gas S.; Barry C.E. III; Tekala F.,
 RA Badcock K.; Basham D.; Brown D.; Chillingworth T.; Connor R.,
 RA Davies R.; Devlin K.; Feltwell T.; Gentles S.; Hamlin N.; Holroyd S.,
 RA Hornsby T.; Jagels K.; Krogh A.; McLean J.; Moule S.; Murphy L.,
 RA Oliver S.; Osborne J.; Quail M.A.; Rajandream M.A.; Rogers J.,
 RA Rutter S.; Seeger K.; Skelton S.; Squares S.; Squares R.,
 RA Sulston J.E.; Taylor K.; Whitehead S.; Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RN Nature 393:537-544(1998).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D.; Alland D.; Eisen J.A.; Carpenter L.; White O.,
 RA Peterson J.; Deboy R.; Dodson R.; Gwinn M.L.; Haft D.; Hickey E.,
 RA Kolonay J.F.; Nelson W.C.; Umayam L.A.; Ermolaeva M.D.; Salzberg S.L.,
 RA Delcher A.; Utterback T.; Weidman J.; Khouri H.; Gill J.; Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Required for binding-protein-mediated phosphate
 CC transport.

CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (Probable).

CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL curation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M30046; AAA25374.1; -
 DR EMBL; Z95209; CAB0484.1; -

DR EMBL; AE006981; AAK45208.1; -
 DR PIR; F70584; F70584.
 DR HSSP; P06128; 11XH.

DR TIGR; MT0961; -
 DR Tuberculin; RV0934; -
 DR InterPro; IPR006059; SBP_bac_1.
 DR Pfam; PF01547; SBP_bac_1; 1.

DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
 KW Antigen; Complete proteome.

FT SIGNAL 1 23 POTENTIAL..
 FT CHAIN 24 374 PHOSPHATE-BINDING PROTEIN 1..
 FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).
 SQ SEQUENCE 374 AA; 38243 MW; 6334968191FF38AA CRC64;

Query Match 36.1%; Score 1831; DB 1; Length 374;
 Best Local Similarity 95.2%; Pred. No. 1.1e-74;
 Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

Qy 59 RIKLEVSFKMRPAQP-----RCGSKPPSPGTPGAGTVAATPPASSVTLAETGSL 112
 Db 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSPGTPGAGTVAATPPASSVTLAETGSL 61

Qy 113 YPLENLWGPAPHERYPNVTTITAGTSGAGTAAAGTAVNIGASDAYLSGDDAAHKGML 172
 Db 62 YPLENLWGPAPHERYPNVTTITAGTSGAGTAAAGTAVNIGASDAYLSGDDAAHKGML 121

Qy 173 NIALAISAQVQVYNLPVSEHLKNGKVLAAVYQGTIKTWDDPQIALNPGVNLPGTAV 232
 Db 122 NIALAISAQVQVYNLPVSEHLKNGKVLAAVYQGTIKTWDDPQIALNPGVNLPGTAV 181

Qy 233 PLHRSDSGDTFLFTQVLSKQDPGKSGPFGTITVDFFPAVPGALGNGGVMVTCGAE 292

Db 182 PLHRSQSGDTFLFTQYLSKQDPGKSGFGFTTVDPAVFGALGNGNGMGWTCACET 241
QY 293 PGCVAYIGISFSDQASQRLGSAQGLNNGSNFLLPDAQSIQAAAAGFASKTPANQAI SWI 352
Db 242 PCVAYIGISFSDQASQRLGSAQGLNNGSNFLLPDAQSIQAAAAGFASKTPANQAI SWI 301
QY 353 DGPADPGYPIINYEVAIVNRRQKDAATATLQAFILHWAITDGNKASFLLDQVHFQPLPPAV 412
Db 302 DGPADPGYPIINYEVAIVNRRQKDAATATLQAFILHWAITDGNKASFLLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
Db 362 VKLSDALIATISS 374

RESULT 3

CF10 MYCTU
ID CF10 MYCTU STANDARD; PRT; 99 AA.
AC O69739;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 10 kDa culture filtrate antigen cfp10 (Secreted antigenic protein
DE MTA-10).
GN CFP10 OR LHP OR WISA10 OR RV3874 OR MT3988 OR MTV027.09.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
RC STRAIN=H37Rv;
RX MEDLINE=99061212; PubMed=9846755;
RX Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,
Gicquel B.;
RA "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
RT low-molecular-mass culture filtrate protein (CFP-10).";
RL Microbiology 144:3195-3203 (1998).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies A., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Biswal W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RA Singh B., Siddiqui Z., Singh S., Sharma P.;
RT "RV3874 (msa-10) gene of a clinical isolate of Mycobacterium
RT tuberculosis from India.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

DR EMBL; AF004671; AAC83445.1; -
DR EMBL; AL022120; CAAL7966.1; -
DR EMBL; AE007190; AAK48356.1; -
DR EMBL; AF149854; AAL14999.1; -
DR PIR; H70802; H70802.
DR TIGR; MT3988; -
DR Tuberculin; RV3874; -
KW Antigen; Complete proteome.
FT INIT MET 0
SQ SEQUENCE 99 AA; 10663 MW; ESCAE6A996C5489D CRC64;

Query Match 9.6%; Score 485; DB 1; Length 99;
Best Local Similarity 100.0%; Pred. No. 9.7e-16;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 AEMKTDATLAQAGNFERISGLDKTQIDQVESTAGSLQCGWGAAGTAAQAAVVRFOEA 485
Db 1 AEMKTDATLAQAGNFERISGLDKTQIDQVESTAGSLQCGWGAAGTAAQAAVVRFOEA 60
QY 486 ANKQKQELDEISTNIRQAGVQYGRADDEQQQALSSQMGF 524
Db 61 ANKQKQELDEISTNIRQAGVQYGRADDEQQQALSSQMGF 99

RESULT 4
PSTS XYLYFA
ID PSTS XYLYFA STANDARD; PRT; 364 AA.
AC Q9PBK3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphate-binding periplasmic protein precursor (PBP).
GN PSTS OR XP2141.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=9845C;
RX MEDLINE=20365717; PubMed=10910347;
RX Simpson A.J.G., Reinach P.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferri J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Honeis J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kurana E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.P., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega P.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

```

RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Vertjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RA "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
CC -!- FUNCTION: REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE
CC TRANSPORT (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE004028; AAF84940.1; -.
DR HSP: G82595; G82595.
DR InterPro: IPR006059; SBP_bac_1.
DR Pfam: PF01547; SBP_bac_1; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Phosphate transport; Transport; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 364 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
SQ SEQUENCE 364 AA; 38736 MW; 8609CFAA159D4277 CRC64;

Query Match 9.1%; Score 463; DB 1; Length 364;
Best Local Similarity 32.8%; Pred. No. 4.1e-14;
Matches 116; Conservative 67; Mismatches 151; Indels 20; Gaps 10;

QY 76 GSKPPSPETGAGATVATPASS-----PVTLAETGSLLYPLFNLWGFAPHERYPNVT 131
DB 21 GCKPSNDQSTGVSDQNGNSTTPPSAEQTKSVKISGAGASFIPLISQWSADYNAATGN-K 79

QY 132 ITAGCTSGAGIAQAAGTNYIGASDAYLSEGDMAHKLMLNIALAISAOQVNNYLPVYS 191
DB 80 INYQISGGGIAQIKATIDFSSDKPLDSEL-TQALGQPSAIGGVVWVNLNLTIE 138

QY 192 E-HKLNKGVLAAMYQGTIKTDDPQIAALNPGVNLPGTAVVPLHRSDGSDTFLFTQYL 250
DB 139 PGKLRATGCLPLADIFLGIKSNDAATISANPGLHLPDTKINTVHRSDGSGTTFNFSNYL 198

QY 251 SKQDPEGWKGSPGTTVDPPAVPGALGNGNGWGTGCAETPGCVAYIGISFLDQASOR 310
DB 199 SKVSAE-WKQKVGEGTSVQW---PGVGKGNGEVSAYVQCIKSGYVELAY---ALQN 251

QY 311 GLGEAQIGNSGNFLPDPAQISQAAA--AGFASKTPANQAISMIDGPADPGYPFIINYEYA 368
DB 252 KMSYTAQNAGAGWQVQPSAESFAAASADWSNAKDFNLVITNATGEA--AWPITATNFI 309

QY 369 IVNNEQKDAATAQLQAFLEHWAIDGNKASFLDQVHFQPLPRAVVKLSDALIAT 422
DB 310 LMRKQTKDAQRKATLDFKWSFENGQKQA--NELHYVPLPPNLVQIOEAYWAS 361

RESULT 5
PSTS RHIL0
ID _PSTS RHIL0 STANDARD; PRT; 351 AA.
AC Q98FL2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphate-binding periplasmic protein precursor (PBP).
GN PSTS OR ML13723
OS Rhizobium loti (Mesorhizobium loti)
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1] ,

```

```

RP SEQUENCE FROM N.A.
RC STRAIN:MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Katanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
CC -!- FUNCTION: REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE
CC TRANSPORT (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP003002; BAB50555.1; -.
DR HSP: P06128; 11XH.
DR InterPro: IPR006059; SBP_bac_1.
DR Pfam: PF01547; SBP_bac_1; 1.
KW Phosphate transport; Transport; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 351 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
SQ SEQUENCE 351 AA; 36766 MW; 2710D1BC0417574D CRC64;

Query Match 8.8%; Score 444; DB 1; Length 351;
Best Local Similarity 34.4%; Pred. No. 2.7e-13;
Matches 112; Conservative 51; Mismatches 149; Indels 14; Gaps 8;

QY 95 TTPASSPVTLAETGSLLYPLFNLWGFAPHERYPNVTITAGCTSGAGIAQAAGTNYIG 154
DB 21 TLSAAIAADLSAGSTFIYPVFAKWADTY-KKDTGVGLNYQISGGGIGKQVIAKTVTFG 79

QY 155 ASDAYLSEGDMAHKLMLNIALAISAOQVNNYLPVSE-HLKNKGVLAAMYQGTIKTWD 213
DB 80 ATDKFMSDADLEKN-GLVQPFPMVGGIVPIVNLTFGKPGELVLDGKTLAQIYLGAITTD 138

QY 214 DPQAALNPGVNLPGTAVVPLHRSDGSDTFLFTQYLSKQDPEGWKGSPGTTVDPPAV 273
DB 139 DAAIKALNPSLTLSTALAVVHRSDGSGTTFNINLVKLSFD-WKDKVGSDDTAVENTP- 196

QY 274 PGALGNGNGWGTGCAETPGCVAYIGISFLDQASQRLGEAQIGNSGNFLPDPAQISQ 333
DB 197 --GVGAKGSEGVANTVKQTDGGIGYVEYAY---AKQNNLSYSKMLNAGKVVESLESFG 251

QY 334 AAAASFASKTPANQAISMIDGPADPGYPFIINYEYAVIIVNNEQKDAATAQLQAFLEHWAID 393
DB 252 AAASNADPKGANKFNVIITNPEGDTTWPIAASTVWLIHKAPDDAATGEALKFFAWAYKD 311

QY 394 GNK-ASFLDQVHFQPLPRAVVKLSDA 418
DB 312 GKETAALDYV--SIPDSVVDLIKA 334

RESULT 6
PSTS ECOLI
ID PSTS ECOLI STANDARD; PRT; 346 AA.
AC P06128; P76744;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphate-binding periplasmic protein precursor (PBP).
GN PSTS OR PHOS OR B3728 OR SF3727.
OS Escherichia coli, and

```


FT STRAND 215 220
 FT HELIX 221 225
 FT TURN 226 228
 FT STRAND 230 231
 FT STRAND 232 235
 FT TURN 233 238
 FT STRAND 237 238
 FT STRAND 241 242
 FT HELIX 246 252
 FT TURN 253 255
 FT STRAND 258 260
 FT STRAND 267 267
 FT TURN 272 273
 FT STRAND 277 286
 FT STRAND 288 288
 FT HELIX 292 308
 FT HELIX 310 315
 FT TURN 316 317
 FT STRAND 319 320
 FT HELIX 323 336
 FT STRAND 338 338
 FT TURN 340 341
 FT STRAND 344 344
 SQ SEQUENCE 346 AA; 37024 MW; 867DA7199C2C87ED CRC64;

 Query Match 8.1%; Score 413; DB 1; Length 346;
 Best Local Similarity 22.2%; Pred. No. 6.4e-12;
 Matches 108; Conservative 56; Mismatches 145; Indels 26; Gaps 10;

 QY 92 TVATTPAS-----SPVTIAETGSTALLYPLFLNLPAPHERYPNVITTAQGTGSGA 141
 DB 7 TVATVVAATLSMAFSVFAEASLTGAGATFPAPVYAKWADTYQKGTGN-KVNYQIGSGG 65

 QY 142 GIQAAGAGTNGASDAYLSEGDMAHKGMLNLTALISAQVNNYLPGV-SEHLKNGKV 200
 DB 66 GVKQITANTVDFGASDAPLSD-EKLAQEGLFQFPFTVIGGVVLAIVNPGLKSELVLDGKT 124

 QY 201 LAAMYQGITKTWDDPQIAALNPGVNLPGTAVVPLHRSDSGDFTFLTYQLSKQDPGNGK 260
 DB 125 LGDIYLGKIKKWDDEATAKLNGLKLPSONIAVVRADSGTGFVTSYLAKNWEE-WKN 183

 QY 261 SPFGITVDPVAPGALGENGNGMVTGCAETPGCVAYIGISFLDQASQRLGEAQLGNS 320
 DB 184 NVGTGSTVKKWPI---GLGGKXNDGIAAFVQRLPGAIGYVEYAY---AKQNNLAYTKLISA 237

 QY 321 SGNFLPDAOSIQAAAAGPA-SKTPANQALSMIDGPDGPGPYPIINYEYALVNNRQKDAAT 379
 DB 238 DGRFVSSTENFANAAGADWSKTFQAQ---DLTNQGEDAWPITSTFFILHKQKKPEQ 294

 QY 380 AQTQLAFLHWAITDGNKASFLDQVHFQPLPPAVVK 414
 DB 295 GTEVLKPFWDWAYKTGAQA--NDLDVASLPDSVVE 327

RESULT 7

ID PSTS PASMU STANDARD; PRT; 344 AA.
 AC QCNJ34;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphate-binding periplasmic protein precursor (PBP).
 GN PSTS OR PM0436.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc., Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC -!- FUNCTION: REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE
 CC TRANSPORT (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE006079; AK02520.1; ALT_INIT.
 DR HSSP; P06128; I1XH.
 DR Pfam; PF01547; SBP_bac_1; 1.
 KW Phosphate transport; Transport; Periplasmic; Signal;
 KW Complete proteome.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 344 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
 SQ SEQUENCE 344 AA; 37362 MW; 6378650112CF9161 CRC64;

 Query Match 8.1%; Score 411.5; DB 1; Length 344;
 Best Local Similarity 31.7%; Pred. No. 7.4e-12;
 Matches 108; Conservative 59; Mismatches 159; Indels 15; Gaps 8;

 QY 92 TVATTPASSPVTIAETGSTALLYPLFLNLPAPHERYPNVITTAQGTGSGAGIAQAAGTV 151
 DB 14 TFAUTTOAQITIGA--GASFPYPIYAKWA-SMVEKQTKGNKNYQISGGGQQIITAKTI 70

 QY 152 NIGASDAYLSEGDMAHKGMLNLTALISAQVNNYLPVSE-HLKLNGKVLAAAMYQGITK 210
 DB 71 DFGASDPMKAEALQAQ-LIQFPAILIGTVPVVNLPEITAGQLKLSGEVLADIFLGKIK 129

 QY 211 TWDDPQIAALNPGVNLPGTAVVPLHRSDSGDFTFLTYQLSKQDPGNGKSPGFTVDF 270
 DB 130 KWNDFAIKLNQGANLPDKAIIVVHRSDSGTFTGNTNYSKVSTE-WKETVQGGSKVK 188

 QY 271 PAVPGALGENGNGMVTGCAETPGCVAYIGISFLDQASQRLGEAQLGNSGNFLPLPDAQ 330
 DB 189 PTGGGKNGEVAAYVSKIKYSGYVEYA-----YAKQNLAWASLONKAGQFVQPSAE 242

 QY 331 SIQAAAAGFASKTPANQALSMIDGPDGPGPYPIINYEYALVNNRQKDAATQTTQAFLHWA 390
 DB 243 SFMAAANAQWESAAGMVGVLITNEEGDTSWPTVAASPIILLHKAKEPTEITKAVDFDFWA 302

 QY 391 ITDGNKASFLDQVHFQPLPPAVVKLSALAT-ISSAEMKT 430
 DB 303 FKQGRVAA--TELDYVPLPEEVIQIQAKWKEVKSDDGKT 341

RESULT 8

ID PSTS HAEIN STANDARD; PRT; 334 AA.
 AC P45192;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphate-binding periplasmic protein precursor (PBP).
 GN PSTS OR H11383.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,


```
RESULT 10
PST3 MYCLE          STANDARD;          PRT;          369 AA.
ID   PST3 MYCLE
AC   Q9CBES;
DT   28-FEB-2003 (Rel. 41, Created)
DI   28-FEB-2003 (Rel. 41, Last sequence update)
DE   Phosphate-binding protein 3 precursor (PBP-3) (Psts-3).
GN   PST2 OR PHOS2 OR ML2095.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- FUNCTION: Required for binding-protein-mediated phosphate
CC   transport (By similarity).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC   (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   EMBL; AL583924; CAC31050.1; -
CC   PIR; B87171; B87171.
CC   HSP; P06128; 11XH.
CC   Leproma; ML2095; -.
CC   InterPro: IPR006059; SBP_bac_1_1.
CC   Pfam: PF01547; SBP_bac_1_1.
CC   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
KW   Complete proteome.
FT   SIGNAL          1..21
FT   CHAIN           22..369
FT   LIPID           22..22
FT   SEQUENCE       369 AA; 37733 MW; 697DECB6A84B9AC CRC64;
SQ
Query Match
Best local similarity 28.9%; Pred. No. 1.9e-08;
Matches 109; Conservative 50; Mismatches 159; Indels 59; Gaps 13;
QY 75 CGSKFPSPGSPETGAGAGTATVPASSPV-----TLAETGSLTLLYFLNLWGPAFHRYVP 128
Db 22 CGS-----DNNAVGSARTGSSCGVSCGGPTLKASGSTAQANAMTFVNAFERSCP 74
QY 129 NVTITAGTSGAGIAQAAGTAVNIGASDAYLSEGDMAHK-----GLMNIALAISAOV 183
Db 75 GQTLNVTANGSAGVSEFNGNQTFDGSDFLSRKEYAAAEQCGSQAWNLPLVFGPIAI 134
QY 184 NYNLPGVSEHLKNGKVLAAVYGTITKTDWDDPOIALNPGVNLPGTAVVPLHRSQSGPT 243
Db 135 TYNVNGUSS-LNLDGPTTKIFNGSIASWMDPAIQALNTGVALPASPIHVHFRNDSGTT 193
QY 244 FLFTQYLSKQDPGWSKSGPGTGTVDPPAVFGALGE--NNGGMVGTGCAETFGCAVYIGI 301
Db 194 DNFORILDVASNGWGK--GIGKT-----FKGGVGBGAKNGDGTSAAVKSTEGSITNEW 246
QY 302 SFLDQASQSGIGLGEAQLGNSGNFLLPDAQSIQRAAAGFASKTFAN-----QAI 349
Db 247 SF---ASARKLNTAKIATSAD---PEPIAISVDSVG---KTISGATIIEGENDLVLDTV 296
QY 350 SMIDGPADPGYPIINYEIVANNROKDAATAQLOALFLHWAITDGNKASFLDQVHFQPLP 409
Db 297 SFYKPAQPGSYPIVLTATYIEVCSKYPAQVGRAVFLQSTIGGQNG--LGDNVYVIP 354
QY 410 PAVVKLSDALIATISSA 426
Db 355 -----DSFKSRUSTA 364
RESULT 11
PST2 MYCTU
ID   PST2 MYCTU          STANDARD;          PRT;          370 AA.
AC   O05870; P96905;
DT   28-FEB-2003 (Rel. 41, Created)
DI   28-FEB-2003 (Rel. 41, Last sequence update)
DE   Phosphate-binding protein 2 precursor (PBP-2) (Psts-2).
GN   PST2 OR RV0932C OR MT0959 OR MTCY08D9.07.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=Erdmann;
RX   MEDLINE=97284492; PubMed=9139906;
RA   Lefevre P., Braibant M., de Wit L., Kalai M., Roeper D.,
RA   Grotzinger J., Delville J.-P., Peirs P., Ooms J., Huygen K.,
RA   Content J.;
RT   "Three different putative phosphate transport receptors are encoded by
RT   the Mycobacterium tuberculosis genome and are present at the surface
RT   of Mycobacterium bovis BCG.";
RL   J. Bacteriol. 179:2900-2906(1996).
RN   [2]
RP   REVISION TO 361.
RA   Content J.;
RA   Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN   [3]
RC   SEQUENCE FROM N.A.
RX   STRAIN=H37RV;
RC   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaita F.,
RA   Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann E.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.B., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   DeCher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC   -!- FUNCTION: Required for binding-protein-mediated phosphate
CC   transport.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
```

```
CC (Probable).
CC -!- INDUCTION: ACCUMULATION OF PROTEIN IS ENHANCED UNDER PHOSPHATE
CC -!- STARVATION.
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48056; CA88137.1; -.
DR EMBL; Z95209; CA808486.1; -.
DR EMBL; AE006981; AAK45202.1; -.
DR PIR; D70584; D70584.
DR HSSP; P06128; 1LXH.
DR TIGR; M0959; -.
DR Tuberculin; RV0932C; -.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
KW Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 370 PHOSPHATE-BINDING PROTEIN 2.
FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 370 AA; 37864 MW; 97F5116CEB9B3B5C CRC64;
Query Match 6.5%; Score 330.5; DB 1; Length 370;
Best Local Similarity 31.1%; Pred. No. 3.1e-08;
Matches 114; Conservative 44; Mismatches 168; Indels 41; Gaps 13;
QY 75 CGSKPPSPGATGAGTAVTTPASSPVTLAETGTLTLLPLFLWGPATHEHYENVITA 134
Dy 23 CGGCTNSS--SGAG-GTSGSVHCGGKXELHSSSTACENAWQFVAYVRSCEYLDY 79
QY 135 QGTSGAGIAQAAGTAVNIGASDAYLS---EGDMAAHK---GLMNTALAIASQVNYNL 187
Dy 80 NANGSGAGVTQFLNNTDFAGSDVPLNPSTQCPDRSAERCGSPAWDLPTVFGPIATYNI 139
QY 188 PGVSEHLKLVLAAMVQGIKTDPOIALNPVNLPGTAVPLHRSDGSDTLFT 247
Dy 140 KGVST-LNLDGPTAKIPNGITVWNPQIQALNSGIDLPTPISVIFRSDXGTSNFO 198
QY 248 QYLSKQDPEGKSGPGFTTVDFFAVPGALGNGMGVTCATPGCVAYIGISFLDQA 307
Dy 199 KYLDGASNGAKG--GASETFNGGVGVGASGNGTSALL---QTDSITNWSF---A 250
QY 308 SQRGLGEAQLGNSGNFLPDAQSIQAAAGFASKTPANQAI-----SMIDGP 355
Dy 251 VGKQLNMAQIITSAG---PDPVAITTESVG---KTIAGAKIMQGGNDLVLDTSFYRPT 303
QY 356 APDGYPIINYEYAVINNRQDAATAQTLQAFHLWAITDGNKASFLDQVHFOPLPPAV-VK 414
Dy 304 QPGSPVILATYEIVCSKYPDATGTAVAFMQAIGPGQEG--LDQVGSIPLEKFSQAK 361
QY 415 LSDALIA 421
Dy 362 LAAAVNA 368
RESULT 12
PST3_MYCTU
ID PST3_MYCTU STANDARD; PRT; 370 AA.
AC O86343; Q50794;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphate-binding protein 3 precursor (PEP-3) (PstS-3) (Antigen Ag88).
GN PST3 OR PHOS2 OR RV0928 OR MT0955 OR MTCY21C12.22.
OS Mycobacterium tuberculosis.
```

```
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RC STRAIN=Brdmann;
RX MEDLINE=97000022; PubMed=8843165;
RA Braibant M., Lefevre P., de Wit L., Ooms J., Peirs P., Huygen K.,
RA Wattiez R., Content J.;
RT Identification of a second Mycobacterium tuberculosis gene cluster
RT encoding proteins of an ABC phosphate transporter."
RL FEBS Lett. 394:206-212(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RC STRAIN=Brdmann;
RX MEDLINE=97284492; PubMed=9139906;
RA Lefevre P., Braibant M., de Wit L., Kalai M., Roepert D.,
RA Groetzig J., Delville J.-P., Peirs P., Ooms J., Huygen K.,
RA Content J.;
RT "Three different putative phosphate transport receptors are encoded by
RT the Mycobacterium tuberculosis genome and are present at the surface
RT of Mycobacterium bovis BCG."
RL J. Bacteriol. 179:2900-2906(1997).
CC -!- FUNCTION: Required for binding-protein-mediated phosphate
CC transport.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -!- INDUCTION: By phosphate starvation.
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95209; CA808486.1; -.
DR EMBL; Z48057; CA88138.1; -.
DR EMBL; AE006981; AAK45202.1; ALT_INIT.
DR PIR; H70583; H70583.
DR HSSP; P06128; 1LXH.
DR TIGR; M0955; -.
DR Tuberculin; RV0928; -.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1.
```

DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
 KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
 KW Antigen; Complete proteome.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 370 PHOSPHATE-BINDING PROTEIN 3.
 FT LIPID 23 370 N-ACYL DIGLYCERIDE (POTENTIAL).
 FT VARIANT 112 114 AAA -> PSQ (IN STRAIN ERDMAN).
 FT VARIANT 116 116 Q -> A (IN STRAIN ERDMAN).
 FT VARIANT 118 118 MISSING (IN STRAIN ERDMAN).
 FT VARIANT 253 254 OH -> HD (IN STRAIN ERDMAN).
 FT SEQUENCE 370 AA; 37953 MW; 7D557829A9A119E0 CRC64;
 SQ SEQUENCE 370 AA; 37953 MW; 7D557829A9A119E0 CRC64;
 Query Match 6.3%; Score 314; DB 1; Length 370;
 Best Local Similarity 30.8%; Pred. No. 1.7e-07;
 Matches 106; Conservative 39; Mismatches 161; Indels 38; Gaps 11;
 QY 86 TGAGAGTATTPASSPV-----TLAETGSTALLPLFLNMGPAFHRYENVNVTITAGGTGS 139
 Db 30 TGGGA---TTGQASAKVDCGKTKLKASGTAQANMTRFVNVEGACPGQTLNTANGS 86
 QY 140 GAGIAQAAGTAVNIGADAYLSEGDMAHK-----GLMNIALAISAAQVYNNLPVGVSEHL 194
 Db 87 GAGISEFNGNOTDFGSDVPLSKDEAAAAQRCCGSPAWNLPVVGPIAVTYNNLSVSS-L 145
 QY 195 KLNGKVLAAVYQTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDSGDTFTFLYLSKQD 254
 Db 146 NLDGPTLAKIFNGSITQWNPALQALNDRFTLFGRIHVFRSDSGTNDNFQRYLQAS 205
 QY 255 PEGWGSFPGTGVTPAVPGALGE--NGNGMVTGCAETPGCAVYIGISFLDQASQRL 312
 Db 206 NGAWGRGAG-----KSGGGVGEAGNDGTSAAAKVTGSIYNEWSF---AAQHL 255
 QY 313 GEALQNSGNFLPDAQIQAAGFASKT---PANQIAISMIDG-----PAPDGPPIY 365
 Db 256 TMANIVTSAGGD--PVAITIDSYGQTIAGATISGVGNDVLVLTDSFYRPREGSPYVILA 313
 QY 366 EYAIYVNRQKDACTAQTALFLHWAITDGNKASFLDQVHFQPLP 409
 Db 314 TYEIVCSKYPDSQVGTAVXAFLOSTIGAGQ--SLGDNNGYIIP 355
 RESULT 13
 RAA3 CHLRE
 ID_RAA3 CHLRE STANDARD; PRT; 1783 AA.
 AC OSPEC4
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Trans-splicing factor Raa3, Chloroplast precursor.
 GN RAA3.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 CC Chlamydomonadaceae; Chlamydomonas.
 CX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=137C / CC-125;
 RX MEDLINE=21161833; PubMed=11285239;
 RA Rivier C., Goldschmidt-Clermont M., Roach J. D.;
 RT Identification of an RNA-protein complex involved in chloroplast
 RT group II intron trans-splicing in Chlamydomonas reinhardtii.;
 RL EMBO J. 20:1765-1773(2001).
 CC -!- FUNCTION: Required for trans-splicing of exons 1 and 2 of the
 CC chloroplast encoded psaA mRNA (a group II intron). May be required
 CC for stability of the chloroplast RNA-protein complex in which it
 CC is found.
 CC -!- SUBUNIT: Part of a 1700 kDa complex that includes the
 CC precursor RNA to exon 1 and the tscA RNA.
 CC -!- SUBCELLULAR LOCATION: Chloroplast stroma.
 CC -!- DOMAIN: The N-terminal 453 amino acids are dispensable, while the
 CC C-terminal 630 amino acids are required for function.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 DR EMBL; AF310675; AAG40000.1; -.
 DR EMBL; AF310674; AAG39999.1; -.
 KW Chloroplast; Transit peptide; mRNA processing; mRNA splicing.
 FT TRANSIT 1 40 CHLOROPLAST (POTENTIAL).
 FT CHAIN 41 1783 TRANS-SPICING FACTOR RAA3.
 FT DOMAIN 112 311 ALA-RICH.
 FT DOMAIN 343 436 SER-RICH.
 FT DOMAIN 478 1078 ALA-RICH.
 FT DOMAIN 1310 1415 ALA-RICH.
 FT DOMAIN 1416 1430 GLN-RICH.
 FT DOMAIN 1496 1506 ARG-RICH.
 FT DOMAIN 1678 1722 ALA-RICH.
 FT DOMAIN 771 778 POLY-ALA.
 FT DOMAIN 920 926 POLY-PRO.
 FT DOMAIN 927 932 POLY-ALA.
 FT DOMAIN 1047 1055 POLY-GLY.
 FT DOMAIN 1318 1325 POLY-ALA.
 FT DOMAIN 1405 1415 POLY-ALA.
 FT DOMAIN 1669 1675 POLY-PRO.
 FT SEQUENCE 1783 AA; 180399 MW; 40F6206BA6BBDDB CRC64;
 Query Match 4.0%; Score 203.5; DB 1; Length 1783;
 Best Local Similarity 20.9%; Pred. No. 0.081;
 Matches 226; Conservative 101; Mismatches 425; Indels 327; Gaps 46;
 QY 52 SAGKITVRIKLEVSFKMRPAQPR-----CGSK-----PSSGPFETG-----AGA 90
 Db 136 SAAPQGTATPAATAKTTPQRPESDADAGSRAQSQYQFGDPGGGALKGAVDAASDAAP 195
 QY 91 GTVATTPASSPVTLAETGSTALLPLFLNMGPAFHRYENVNVTITAGTGSAGATAAAAGT 150
 Db 196 DVAASAP--PAGISDQLSTPACP-----PEREPQ---AGKPRASGRAFAAGVGP 241
 QY 151 VNIGASDAYLSEGDMAHKGLMNIALAISAAQVYNNLPVGVSEHLKNGKVLAAVYQTIK 210
 Db 242 QDVGSGGCACAPAPDESHMGL-----THRDQGHDERIS-----QTAGE 279
 QY 211 TWDDPQIAA--LMPGVNLPGTAVVP-----LHRSDSGDTFTFLYLSKQDPEGWG 259
 Db 280 AWKAGAVAAAPPAATPPSPGLAAAPTFLASSALGTHSSDGD-----MRAVP---G 327
 QY 260 KSPGFGTVDFFPAVPGAL-----GENGGMVTGCAETPGCV---AYIGISFLD- 305
 Db 328 RD-----TPSLSAVAGPVTLGSSSSSSSSSSSSNTSTSTSTNGVTITSNVGVANGSP 382
 QY 306 ----QASQRLGEAQ-----LGNSSGNFL-LPDQASQIAAAGFASKTAPANQIAISMIDGPA 356
 Db 383 QERLMAAFRAVVTQWNTHLGRGRGFAPLPTGGMSTATSAASSSTSSASSSSMDG-- 440
 QY 357 PDGYPIINYEYAVIYVNRQKDAATAQTALFLHWAITDGNKASFLDQVHFQPLP- 409
 Db 441 -----SNAKKTSDAVSLPV-----GQCPAAEQPHVPTAFGSPSQTG 477
 QY 410 -----PAVVKLSDALIATISSAEMTKDAAATLAQEAQNERISG----- 447
 Db 478 ASAVAAQAPSSAMPTAAMAATMGSAATLPTAAVSVSSAAAEQTGSPGLLAGRPA 537
 QY 448 ----DLATQTDQVETAGSLQGWGAAGTAQAAAVVRFQEAANKQKQELDELSTNIRQA 503
 Db 538 LLGRTIQRARIQAAREALRAARHARVGAAMPVPV-----CARPVQSGSQVPPV 589
 QY 504 GVQYSRADEBQQALSSQMGFTQSQTVTVDQBIINRANEVEAPMADPPFDVPTCELT 563
 Db 590 G-----QGVQSOPGRROEPAASATK---LHVADGLPARPVQPAVSAIDLQDTA 636
 QY 564 AAKNAOQLVLSADNMEYLAAGAKERQRLATSLRNAA-KAYGEVDEEA--ATALONDGE 620


```

QY 448 DLKTDQID--QVESTAGSLQGWGAGTA-----AQAAVVFQEAANKQKQELDEIST 498
Db 518 IVEVITDGRKXGSLVGAEEVRSGGIGMENLMEYAVSTMILINWLDVAPENDL-QIRC 576
QY 499 NIR----QAGVQ--YGRADPEQQALSSQM-GFTQSTVTVDQOBILNR-----ANEVEA 546
Db 577 HIRAQFISGICKRLLSKMGFQVEVDKQIEHFRENE--AIDVEDLLQRESSTKDSIEG 634
QY 547 PMADPTDVPITPCETAA-----KNAAQOLVLSADN----- 578
Db 635 EVKD-WTD---PLQITDAIAGRLNTRAHDFVLSALQHLIRNSGDBGRLMYQLVDA 689
QY 579 MEYLAAGAK-----BRQLATSLRNAKAYGEVDEEATAALDNDGEGTVQAESAGA--- 630
Db 690 MLSYVAMDRRLPLDLRQGLTFVQSLDLRL-HTDAEARRAYDESLEARQIAEAALAEED 748
QY 631 -----VGDGS-----SAELTDTPRV-ATAGEFNFMDL 656
Db 749 EMKAQVELGADGLVRKLOKQIEBQTGIIELQSRQNMELKAEALADYQRLRAQELQENELET 808
QY 657 KEA-----ARKLETGDCQASIAHFADGWNFTNL-----TLQ 687
Db 809 RELYMLRDAQDIAASNAKSNNGEAEETPAHMRGILDRKLLTLREKOLERTKTQFKLE 868
QY 688 GDV-----KRGRF-DNWEGDAATACEASLDQQRQWILHWAKLSAAMAKQAQYVAQLH 739
Db 869 GKWVGCHDPSDLRELREQMGDAGP--REAFEEQARLNLSLNPVGSVVRKKT-YIQGME 925
QY 740 VVARREHPTYED-----IVGLER-----LYAENP-----SARDQILPV 772
Db 926 DTAEBELQTDDEVVAKARLVLDLRPRMDPQATGLLGEIAAKVKIDADDADKDEGRP- 984
QY 773 YAEYQQRSEKVLTEYN-----NKAALBFVNPKPPPAI-----KIDPPPPPBQGG 817
Db 985 -TESEQPAEGAAATKDEQGVDDT-VAVDKATAAPPPPPPPPAHPLGSLGAAPPPPPPP- 1041
QY 818 LIPGLFMPSPDSGVTPTGTMAPAAVMVPTGSPGGLPADTAQLTSAGREAAALSGDVA 877
Db 1042 -----PPPGGAGAAPPP-PPPPPPPPPGGLGGPPPPPPPP----- 1076
QY 878 VKAASLGGGGGGVSPAPLGSIGAESVRPAGAGDIAGLQGRAGGGAALGG 930
Db 1077 -----PGGFGGPPPPPPPGGFGGPPPPPPPGGAFGVPPPPPPPGTVIGG 1123

```

RESULT 15

```

CF10_MYCLE
ID CF10_MYCLE STANDARD; PRT; 99 AA.
AC O33084;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 10 kDa culture filtrate antigen cfp10 homolog.
GN ML0050 OR MLCB628.13C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Dutboy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy J., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,

```

```

RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.; decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMEL; Y14967; CAA75210.1; -.
DR EMEL; AL583917; CAC29558.1; -.
DR PIR; T10032; T10032.
DR Leproma; ML0050; -.
KW Complete proteome.
FT INIT MET 0
SQ SEQUENCE 99 AA; 10833 MW; 859B484F7EFESA8A CRC64;
Query Match 3.6%; Score 184; DB 1; Length 99;
Best Local Similarity 39.4%; Pred. No. 0.021;
Matches 39; Conservative 20; Mismatches 40; Indels 0; Gaps 0;
QY 426 AEMKTDATLAQAGNPERISGLDKTKIQDQVESTAGSLQGWGAGTAQAQAVVFQEA 485
Db 1 AEMITEAAILTQQAQFDQIASGLSQERNFVDSIGSFQNTWEGQAASALGALGRFDEA 60
QY 486 ANKQKQELDEISTNIRQAGVQYGRADPEQQALSSQMGP 524
Db 61 MQDQIRQLESIVDKLARSGGNYTKTDDANQLLSKQNF 99

```

Search completed: November 21, 2003, 16:04:24

Job time : 13:5909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 48.3698 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-54

Perfect score: 5072

Sequence: 1 MGHHHHHHHVILIGTSPSTW.....PANTEAVIGNRRRODSKESK 983

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1552.5	30.6	374	2 Q49590	Q49590 mycobacteri
2	1412	27.8	374	2 Q49589	Q49589 mycobacteri
3	483.5	9.5	363	16 Q8PAF7	Q8PAF7 xanthomonas
4	482.5	9.5	392	16 Q8VND0	Q8VND0 anabaena sp
5	469.5	9.3	363	16 Q8PM56	Q8PM56 xanthomonas
6	460.5	9.1	373	17 Q8YGS6	Q8YGS6 aeropyrum p
7	456	9.0	347	16 Q8YD9	Q8YD9 anabaena sp
8	456	9.0	375	17 Q8U247	Q8U247 pyrococcus
9	452	8.9	336	2 Q49675	Q49675 mycobacteri
10	445	8.8	403	16 Q8RY26	Q8RY26 deincococcus
11	444	8.8	383	16 Q55199	Q55199 synechocyst
12	440	8.7	346	16 Q8Z993	Q8Z993 yersinia pe
13	438	8.6	339	16 Q8PM55	Q8PM55 xanthomonas
14	435	8.6	339	2 Q56833	Q56833 xanthomonas
15	431	8.5	346	2 Q8KY90	Q8KY90 edwardsiell
16	425	8.4	314	2 Q9AML6	Q9AML6 edwardsiell

ALIGNMENTS

RESULT 1

Q49590 ID Q49590 PRELIMINARY; PRT; 374 AA.
AC Q49590;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE P8CS subunit of ABC transporter.
GN P8CS1B.
OS Mycobacterium intracellulare.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35761;
RX MEDLINE=97055782; PubMed=8900068;
RA Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,
RA Moreno C., Ivanyi J.,
RT "Duplication of genes encoding the immunodominant 38 kDa antigen in
Myobacterium intracellulare.";
RL FEMS Microbiol. Lett. 144:235-240(1996).
DR EMBL; X95538; CAA64784.1; -.
DR HSP; P06128; IIXG.
DR InterPro; IPR006059; SBP_bac.1.
DR Pfam; PF01547; SBP_bac.1_1.
SQ SEQUENCE 374 AA; 38088 MW; A7F2D5843860B9D7 CRC64;

Query Match 30.6%; Score 1552.5; DB 2; Length 374;
Best Local Similarity 83.1%; Pred. No. 5.6e-71;
Matches 294; Conservative 25; Mismatches 34; Indels 1; Gaps 1;
Qy 71 AQPRCGSKPPSGPETGAGTATVPASSPVTLAETGSTLLYPLNMGPAFHPYPNV 130
Db 20 AAACGNSQTGAPSQ-GGGPVATTPASSPVTLSGTSTLLYPLNMGPAFHDKISNV 78
Qy 131 TTTAQGTGSGAGIAQAAGTWNIGASDAYLSEGDMAHKLGMNIALAISAQVNYNLPGV 190
Db 79 TTTTQGTGSGTGI-SQAAAGAVAGISDAYLSEGDMAHKLGMNIALAISAQVNYNLPGV 138

Qy	191	SEHLKNGKVLAAMYGGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDGSGDTFLFTQYL	250
Dd	139	TEHILKNGKVLAMGYNGSVKWNDDPQIAGLNPVNLPATFVPLHRSDGSGDTFLFTQYL	198
Qy	251	SKQDPBGWKSPGFGTITVDFFPAVPGLALGENGGMTGCETPCGVAYIGISFLDQASOR	310
Dd	199	SKQDPBGWKSPGFGTITVAFTVPGLALGENGGMTGCCADTPCGVAYIGISFLDAQOGK	258
Qy	311	GLGEAQLGNSSGNFLPDRAQSIOAAAAGFASKTPANOAI SMIDGPAPDGPIINYYEYIV	370
Dd	259	GLGEAQLANASDKYLLPDAKSIOAAAAGFASKTPANOAILSLINGPADPGPIVNYEYIV	318
Qy	371	NNRQXDAATAQTLOAFLEHWAITDGNKASFLDQVHFQPLPAAVKLSDALIATIS	424
Dd	319	NSGQXDAATAQTLOAFLEHWAISDGNSSSFLDKVHFQPLPADVAKLSAQIAKIS	372
 RESULT 2 Q49589 PRELIMINARY; PRT; 374 AA.			
ID	Q49589;		
AC	AQ589;		
CD	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DI	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	PstS subunit of ABC transporter.		
DS	PST31A.		
OS	Mycobacterium intracellulare.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OX	Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
NCBI_TaxID=1767;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 35761;		
RX	MEDLINE=97055782; PubMed=8900068;		
RA	Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,		
RT	Moreno C., Ivanyi J.		
RT	"Duplication of genes encoding the immunodominant 38 kDa antigen in		
RL	Mycobacterium intracellulare.";		
RR	FEMS Microbiol. Lett. 144:235-240 (1996).		
DR	EMBL; X95538; CAAG64783.1; -		
DR	HSSP; P06128; IA54.		
DR	InterPro; IPR006059; SBP bac_1.		
DR	Pfam; PF01547; SBP bac_1; 1.		
SQ	SEQUENCE 374 AA; 37998 MW; 45796D4B9F6F513D CRC64;		
 Query Match 27.8%; Score 1412; DB 2; Length 374; Best Local Similarity 76.5%; Pred. No. 7.2e-64; Matches 267; Conservative 33; Mismatches 49; Indels 0; Gaps 0			
Qy	75	CSGKPSPSPETGAGAGTVATTPTASSPVTLAETGSTLLYPLPLNWGPAPHFPNVTTFA	134
Dd	24	CGAHATRGAPSGPPAGSVATPATSIIEISEGSTLLYPLPSFGAAYARKHPNVTVT	83
Qy	135	QQTGSAGIAQAACATVNIIGASDAYLSEGDMAAHKGLMNTALAISAQQVNNYLPVGSEHL	194
Dd	84	QSGSGAGISQAAGAAGVIGASDAYLSAGDMAAHHKGLMNTALAISAQQVNNYLPVGHQL	143
Qy	195	KLNGKVLAAMYGGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDGSGDTFLFTQYLSKD	254
Dd	144	KLNGKVLAMGYGTIKTWNDPQIAGINPGQLPATFVPLHRSDGSGSTFOFTQYLSKD	203
Qy	255	PSGWKSPGFGTITVDFFPAVPGLALGENGGMTGCETPCGVAYIGISFLDQASORGJGE	314
Dd	204	PDWGGRSPGFGTITVFPAVPALGENGDGMVTTCGAANPGCVAYTIGSIDAEADQHGJGE	263
Qy	315	AOLGNSSGNFLPDRAQSIOAAAAGFASKTPANOAI SMIDGPAPDGPIINYYEYIVNRQ	374
Dd	264	AKLGNASGKYLLPNKAKSIOAAAAGFASKTPANOAILSLINGPATDGPPIVNYEYIVNADQ	323
Qy	375	KDAATAQTLOAFLEHWAITDGNKASFLDQVHFQPLPAAVKLSDALIATI	423
Dd	324	RDATAQTLOAFLEHWAITEGNGPSFLDKVHFQPLPAEVVTLSDAQIAKI	372

CC Xanthomonadaceae; Xanthomonas.
 ON NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=2022145; PubMed=12024217;
 RA da Silva A.C.R., Perro J.A., Reinach F.C., Parah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Canavan F., Cardozo J., Chamberg F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Pormighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Nadeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Tindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities."
 RL Nature 417:459-463 (2002).
 RL ENBL; AB011790; AAM36445.1; -
 DR InterPro; IPR006059; SBP_bac_1.
 DR Pfam; PF01547; SBP_bac_17.1.
 KW Complete proteome.
 DR Complete proteome.
 KW Complete proteome.
 SQ SEQUENCE 363 AA; 37641 MW; 1A39C0289BF7D3DF CRC64;
 Query Match 9.3%; Score 469.5; DB 16; Length 363;
 Best Local Similarity 32.2%; Pred. No. 3.5e-16;
 Matches 119; Conservative 69; Mismatches 157; Indels 25; Gaps 11;
 QY 52 SACKIVRIKLEVSFKMRPAQRCGSKPPSGSPETGAGAGTAVTTPASSPVTLAETGSL 111
 Db 9 TALSLLIALAACSCKDAQSADA-K---GAPAGAAAD-----SKGAEISGAGASF 59
 QY 112 LYPFLNLWGPAPHERYPNVITTAQGTGSGAGIAQAAGTYNIGASDAYLSEGDMAAHKGL 171
 Db 60 IYPLVSKWSADYNAATGN-KVNYQSIGSGGGIAQIKAGTVDFGSTDKPLDSAE-L-QQAGL 117
 QY 172 MNTALISAQQVYNNIPGVSE-HKLNKVKLAAMYOGTITKTDPPQTAALNPGVNLPGTA 230
 Db 118 GQPSAIGGVVVPVNLGPIAGPKRLITGALGIDIFLGKVTMMNDAAALVAANPGVTLPATK 177
 QY 231 WPLHRSDSGDFTLFTQYILSKODPEGWKSPQGTITVDFPVPALGNGNGMVGTCGA 290
 Db 178 INLVHRSDSGGTTTFNSYLSKVSP-EWKSXVGEISVQW---PGGVGKNGEGVASVYQ 233
 QY 291 ETGCVAYIGISFLDQASQKGLGEAQLGNSGNGFLPLDPAOSIOAAAAG--FASKTPANQA 348
 Db 234 QIKSGISGYVELAY---ALQNKMPYTSIQNAAGOWIEPNAESFRAAASADAWANARDFNLV 290
 QY 349 ISMIDGPAFGYPIINVEYAVNNRKDAATAOTLQAFHLWATDGNKAGFLDOVHFOPL 408
 Db 291 IT--NAPGEKAWITATNFMFMHKQPKDARSKATLDFKWLENGQAQA---SELHYVPL 346
 QY 409 PPAVVKLSDA 418
 Db 347 PPELVKQIEA 356
 RESULT 6
 ID Q9YG56 PRELIMINARY; PRT; 373 AA.
 AC Q9YG56
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 373AA long hypotheical phosphate-binding periplasmic protein.
 GN APE0045.
 OS Aeropyrum pernix.

QY 139 SGAGTAQAAAGTGNIGASD-----AYLSEGDMAAHKGLMNLIALISAQOQVNNLPG 189
Db 78 SHGQEAFLKGLTIDGRTPPVKESTWKFLTGQDP-----LQFPVVGAVVWHNPN 132
QY 190 VSEHLKNGKVLAAWYQGTITKTDPOIAALNPGVNLPGTAVVPLHRSDDSGDTFLFTQY 249
Db 133 VDE-LKLDGETLAKIFMGQIEYDDPAIKSLNPNVNLPHKEIIIVHRSDDSGGTATFTTY 191
QY 250 LSKDPPEGWGKSPGGTTVDPPA--VEGALGNGNGWVTGCASTPCGVAYIGISFLDQA 307
Db 192 LSLVSEE-WATKVAGKTVDPVGVKGIQAKNGPVQQLKNTFISIAYTELSF---A 247
QY 308 SQRLGELAQNGSGNFFLLPDAQSIQAAAAGFASKTP-----ANQAISMIDGPADPGYP 361
Db 248 IEEENLKI VALKXKAGNFVKPTEETIKAAVAAVKAFIPDPAEGYKEDIRQLLNAPGENSY 307
QY 362 LIINVEYATV---NNRQKDAATQTLQAFLHWAITDGNKASFLDQVHFQPLPRAVVKL 415
Db 308 IVAFTHFLVWQNGKGHYSKEKAQIKDFLRWLTEGQKPENI-APGVGLPPEVAEI 364

RESULT 9
Q49675
ID Q49675 PRELIMINARY; PRT; 336 AA.
AC Q49675;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PabB.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96084954; PubMed=7489918;
RA Doukhan L., Predich M., Nair G., Dussurget O., Mandic-Mulec I.,
RA Cole S.T., Smith D.R., Smith I.;
RT "Genomic organization of the mycobacterial sigma gene cluster.";
RL Gene 165:67-70(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBSJ databases.
DR EMBL; J00012; AAA85925.1; -;
SQ SEQUENCE 336 AA; 35916 MW; 0C7CABDB3CBD99F CRC64;

Query Match 8.9%; Score 452; DB 2; Length 336;
Best Local Similarity 40.1%; Pred. No. 2.4e-15;
Matches 122; Conservative 22; Mismatches 74; Indels 86; Gaps 11;

QY 93 VATTPASSPVTLAETGSLTLLYPLFNWGP-AFHERYPNVTTITAGTSGAGIAQAAGTV 151
Db 39 VTTATATSKVTLSKTGSTLNLPLNLSYLLRSKYPNVTTITQGT---GIAQAAGTV 94
QY 152 NIGASDAYLSEGDMA-AHK-----GLMNLIALISAQOQVNNLPGVSEHLKNGKVLAAWYQ 206
Db 95 DIGVSDAYFTVGRHNSNAHKIDHERYSHLSANQLGSVRHQAPQTEWQAG-----AAATQ 150
QY 207 GTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDDSGDTFLFTQY-----LFTOYLSKQPEGNG 259
Db 151 GTIKTWNDPQITALNPGVNLPGTAVVPLHRSDDSGDNTLVHLTHYLAQPDGNGISRL 210
QY 260 -----KSPGF-----GTTVDPPAVPGALGNGNGWVTGCAET 292
Db 211 ASARPPTSPSFAHWAKQQRWHSYQLRLOQASTSTPTRRGGARANRAMPVNTCCST 270
QY 293 PGCVAYIGISFLDQASQRLGELAQNGSGNFFLLPDAQSIQAAAAGFASKTP-ANQAISM 351
Db 271 PKV-----FRLQPAS-----PRTPRRSKVTHL 293
QY 352 IDGP 355
|||

Db 294 IDGP 297
RESULT 10
Q9RYZ6
ID Q9RYZ6 PRELIMINARY; PRT; 403 AA.
AC Q9RYZ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Phosphate ABC transporter, periplasmic phosphate-binding protein.
GN DRA0157.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001862; AAF12207.1; -;
DR HSP; P06128; IXG.
DR TIGR; DRA0157; -;
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1; 1.
KW Complete proteome.
SQ SEQUENCE 403 AA; 42060 MW; 38DC61C6E7D290D0 CRC64;

Query Match 8.8%; Score 445; DB 16; Length 403;
Best Local Similarity 31.5%; Pred. No. 6.9e-15;
Matches 119; Conservative 60; Mismatches 157; Indels 40; Gaps 11;

QY 73 PRCGSKPPSGS-----PETGAGAGT---VATTPASSPVTLAETGSLTLLY 114
Db 40 PLCGTSFVFGSLDAPLTKPHTERRMKKTLGLSALVMISTAAQAQAITGA--GASFPYP 97
QY 115 LFNLMGPAPHERYPNVTTITAGTSGAGIAQAAGTVNIGASDAYLSEGDMAAHKGLMN 173
Db 98 LYS-----KMFSEYKASNVNYQSVSGSGQKILERTVDFAGSDNPNMTDAQLGSAPGTL 153
QY 174 IALAIQAQVNNLPGVSEHLKNGKVLAAWYQGTITKTDPOIAALNPGVNLPGTAVVP 233
Db 154 VPTAIGAVVPAYVLPVGTPLNFDGPTLANIYLGKIKTWGDDPAIKLNPGVTIFPLPITV 213
QY 234 LHRSDSGDGTFLFTQYLSKQDPPEGWKGSPFGTVDPPAVPGALGNGNGWVTGCAETP 293
Db 214 ARSDSGDGTTFVFDYLSKVSGB-WKSKVGAGNSLQWPGVGTGAKGNDGVAGVVGK--TP 269
QY 294 GCVAYIGISFLDQASQRLGELAQNGSGNFFLLPDAQSIQAAAAGFASKTPANQAISMID 353
Db 270 GAGYVELVY---AKQNKJSGFAGVKNRAGKFIADNGNPASNAALGV--IPADTRVSLTN 324
QY 354 GPAPDGPDIINVEYAVIWNRRK---DAATAQTLQAFLHWAITDGNKASFLDQVHFQPLP 409
Db 325 SANAGAPIASFTYLFYKDKYGNRTAQAKALLTVTVVTSQQ--YNEGLDYAKLP 382
QY 410 PAVVKLSDALIATIS 424
Db 383 SNVAAKAKTIINSMN 397

RESULT 11
Q55199


```

RESULT 13
Q8PM55 PRELIMINARY; PRT; 339 AA.
ID Q8PM55
AC Q8PM55
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Phosphate binding protein.
GN PHOX OR XAC1578.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fotmighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011790; AAM36446.1; --
DR InterPro; IPR005673; Peri-phosph.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1; 1.
DR TIGRFAMs; TIGR00975; 3a0107s03; 1.
XW Complete proteome.
ZK
SQ SEQUENCE 339 AA; 35623 MW; 90E16808A4A710B1 CRC64;

Query Match 8.6%; Score 438; DB 16; Length 339;
Best Local Similarity 32.1%; Pred. No. 1.2e-14;
Matches 114; Conservative 60; Mismatches 153; Indels 28; Gaps 9;

QY 65 SFKRPAPQRCGSKPPSPETGACGATVATPASPVTLAETGTLTLPFLNLGPAFH 124
DB 4 SFKTRLA-----VGVLAAALCAQAADVTGA--GASFTYPMVSKWS-ADY 46

QY 125 ERYPNVTITTAQGTSGAGIAQAAGTIVNIGASDYLSEGDMAAHKGLMNLALISAQV 184
DB 47 NAATKKQVNYQSIGSGGIAQIKAAVSDFGSSDAPLKPEELAA-AGLAQFSPVIGVVPV 105

QY 185 YNLPGVSE-HLKLNGKVLAAVQGTIKTWDPPQIAALNPGVNLPGTAVVPLHRS DSGSDT 243
DB 106 VNVPGVAAAGALKLDGKTLGDIFLGKYSTWNDPAIALNPGVKLPBGKITVHRS DSGSGT 165

QY 244 FLFTQYLSKQDPGEGKSPGFGTTVDPAVPGALGNGMGWVTGCAETPGCVAYIGISF 303
DB 166 FNTYLSKVNPD-WGKVGEGTAVQWPT---GIGGKNEGVAAYVKQIKGGIGVVELSY 221

QY 304 LDQASQRLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAISMIDGPAPGVPII 363
DB 222 ---ALQNKVAYTAMKNAAGKVFQPSDETFAAANSADWGSSKDFYLVMTNAAGDNWPI 278

QY 364 NYEYAIYNNRQKDAATAQTQLQAFHLWAITDGNKASFLDQVHFQPLPPAVVKLSDAL 422
DB 279 ATNFIIVQKKPKNPAGLKNLTLEFFRWYTKGDAQA--KOLDYVPLPDALVTQIEAY 335

RESULT 15
Q8KY90 PRELIMINARY; PRT; 346 AA.
ID Q8KY90
AC Q8KY90
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Phosphate-binding protein.
GN PSTs.
OS Edwardsiella tarda.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Edwardsiella.
OX NCBI_TaxID=636;
RN [1]
RP SEQUENCE FROM N.A.
RA Tan Y.P., Leung K.Y.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 32.6333 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672a-64

Perfect score: 2737

Sequence: 1 MHHHHHTAASDNFQLSQGG.....RAPVEADAGGQKVLVRNV 539

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

-Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2737	100.0	543	22 AAU01905	M. tuberculosis an
2	1963	71.7	392	20 AAU73765	M. tuberculosis an
3	1963	71.7	392	20 AAU73555	M. tuberculosis an
4	1963	71.7	392	22 AAU08226	Polypeptide encode
5	1963	71.7	392	22 AAU01882	M. tuberculosis an
6	1963	71.7	392	23 AAU05959	M. tuberculosis an
7	1963	71.7	392	23 AAU05959	M. tuberculosis an
8	1963	71.7	392	23 AAU05959	Mycobacterium sp.
9	1941	70.9	408	22 AAU01886	Mycobacterium spec

10	1941	70.9	408	22 AAU01887	M. tuberculosis an
11	1718	62.8	788	22 AAU01903	M. tuberculosis an
12	1263	46.1	394	23 AAU05402	M. tuberculosis an
13	1168	42.7	242	22 AAU01883	M. tuberculosis an
14	1060	38.7	744	22 AAU01902	M. tuberculosis an
15	1060	38.7	815	22 AAU01904	M. tuberculosis an
16	1055	38.5	219	22 AAU01884	M. tuberculosis an
17	809.5	29.6	729	20 AAU32059	Mycobacterium tube
18	809	29.6	726	23 AAU74588	Antigenic fusion p
19	809	29.6	729	22 AAU02142	Ra12-H9-32A fusion
20	809	29.6	729	23 AAU02142	Mycobacterium sp.
21	809	29.6	729	23 AAU02142	Mycobacterium spec
22	809	29.6	729	23 AAU02142	Mycobacterium spec
23	809	29.6	729	23 AAU02142	Mycobacterium spec
24	809	29.6	729	23 AAU02142	Mycobacterium spec
25	765	28.0	577	22 AAU02140	Ra12-W11 fusion pr
26	749.5	27.4	220	22 AAU02141	Ra12-mammaglobin f
27	739	27.0	231	20 AAU32071	Mycobacterium tube
28	739	27.0	234	23 AAU74600	Antigenic fusion p
29	737.5	26.9	585	22 AAU83277	Chlamydia trachoma
30	737.5	26.9	585	23 AAU83277	Chlamydia trachoma
31	728	26.6	583	22 AAU83281	Chlamydia trachoma
32	728	26.6	583	23 AAU83281	Chlamydia trachoma
33	727	26.6	230	22 AAU83281	Chlamydia trachoma
34	722.5	26.4	518	22 AAU83276	Chlamydia trachoma
35	722.5	26.4	518	23 AAU83276	Chlamydia trachoma
36	718.5	26.3	529	23 AAU83276	Breast tumour-spec
37	717.5	26.2	525	21 AAU83276	C. pneumoniae sero
38	717.5	26.2	525	22 AAU83276	Protein encoded by
39	717.5	26.2	525	23 AAU83276	Chlamydia trachoma
40	709.5	25.9	715	22 AAU83273	Chlamydia trachoma
41	709.5	25.9	715	23 AAU83273	Chlamydia trachoma
42	708	25.9	715	22 AAU83275	Chlamydia trachoma
43	708	25.9	715	23 AAU83275	Chlamydia trachoma
44	706.5	25.8	384	23 AAU85596	Ra12-L985P fusion
45	706.5	25.8	384	24 AAU69568	Human Ra12-L985P f

ALIGNMENTS

RESULT 1

AAU01905

ID AAU01905 standard; Protein; 543 AA.

XX AC AAU01905;

XX DT 29-AUG-2001 (first entry)

XX DE M. tuberculosis antigen HTCC#1 fusion protein #4.

XX KW TbrA12-HTCC#1; antigen; vaccine; tuberculosis;

XX KW AIDS; acquired immunodeficiency disease; His Tag.

XX OS Mycobacterium tuberculosis.

XX OS Synthetic.

XX FH Key

XX FT Binding-site

XX FT Location/Qualifiers

XX FT 3..8

XX FT /label= Histidine tag

XX FT /note= "Nickel chelating region used to aid

XX FT purification of the protein"

XX FT /label= OTHER

XX FT /note= "In frame STOP codon"

XX FT Misc-difference 541

XX FT WO2000124820-A1.

XX PN 12-APR-2001.

XX PD 10-OCT-2000; 2000WO-US28095.

XX PF 07-OCT-1999; 99US-0158338.

XX PR

PR 07-OCT-1999; 99US-0158425.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 PI WPI; 2001-290576/30.
 DR
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 XX
 XX Claim 17; Fig 8; 168pp; English.
 XX
 CC The sequence represents Mycobacterium tuberculosis fusion protein,
 CC TbrA12-HTCC#1 and includes a His tag at the N-terminus to aid
 CC purification. Compositions comprising at least 2 heterologous
 CC antigens, as a fusion protein, and vectors expressing the fusion
 CC proteins, are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS.
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 2737; DB 22; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.2e-219; Indels 0; Gaps 0;
 Matches 539; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHTAASDNFQSGGQGFAPICQAMAIAGQIRSGGSPVHIGPTAFILGLGVD 60
 Db 2 MHHHHTAASDNFQSGGQGFAPICQAMAIAGQIRSGGSPVHIGPTAFILGLGVD 61
 QY 61 NNGGARVORVVGSPAPASLGISTGDVITAVDGPINSATAMALNGHHGPDVISTWQ 120
 Db 62 NNGGARVORVVGSPAPASLGISTGDVITAVDGPINSATAMALNGHHGPDVISTWQ 121
 QY 121 TKSGGTRTGNVTLAEGPPAEFLVPRGMSRAFIIDPTISAIDGLYDLIGIPNQGGLY 180
 Db 122 TKSGGTRTGNVTLAEGPPAEFLVPRGMSRAFIIDPTISAIDGLYDLIGIPNQGGLY 181
 QY 181 SLEFFFEKALEELAAFPDGLWLSAADKYAGKRNHNVFFQELADLDRLQISLIHQAN 240
 Db 182 SLEFFFEKALEELAAFPDGLWLSAADKYAGKRNHNVFFQELADLDRLQISLIHQAN 241
 QY 241 AVQTRDILEGAKGLEFVRPVAVDLTYIPVVGHALSAFQAPFCAGAMAVVGGALAYLV 300
 Db 242 AVQTRDILEGAKGLEFVRPVAVDLTYIPVVGHALSAFQAPFCAGAMAVVGGALAYLV 301
 QY 301 VKTLINATOLLKLLAKLAELVAAATADIISDVADIIGKILGEVWEFITTALNGKELWDX 360
 Db 302 VKTLINATOLLKLLAKLAELVAAATADIISDVADIIGKILGEVWEFITTALNGKELWDX 361
 QY 361 LTGWVTGLFSRWGNSNLESFFAGVPLTGLTATSLGQVTLFGAAGLSASSGLAHADSLS 420
 Db 362 LTGWVTGLFSRWGNSNLESFFAGVPLTGLTATSLGQVTLFGAAGLSASSGLAHADSLS 421
 QY 421 ASLPALAGIGGSGFGGLPFLSAQVHAASTROALPRADGPVGAAGAAQVGGOSQLVSAQGS 480
 Db 422 ASLPALAGIGGSGFGGLPFLSAQVHAASTROALPRADGPVGAAGAAQVGGOSQLVSAQGS 481
 QY 481 QMGGPGVGMGNHPSGASKGTTTKYSEGAAGTDAERAPVADAGGQKVLVRNV 539
 Db 482 QMGGPGVGMGNHPSGASKGTTTKYSEGAAGTDAERAPVADAGGQKVLVRNV 540
 RESULT 2
 ID AAW73765
 XX AAW73765 standard; Protein; 392 AA.
 AC AAW73765;
 XX

24-MAR-1999 (first entry)
 M. tuberculosis antigen clone hTcc#1 protein sequence.
 Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
 infection.
 Mycobacterium tuberculosis.
 WO9853075-A2.
 26-NOV-1998.
 20-MAY-1998; 98WO-US10407.
 05-MAY-1998; 98US-0073010.
 20-MAY-1997; 97US-0859381.
 (CORI-) CORIXA CORP.
 Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAW;
 WPI; 1999-045314/04.
 N-PSDB; AAX01177.
 Polypeptide comprising immunogenic Mycobacterium tuberculosis
 antigen - useful for immunisation against M. tuberculosis infection
 to treat or prevent tuberculosis, and in diagnosis of tuberculosis
 Claim 2; Page 87-88; 100pp; English.
 This sequence represents an immunogenic portion of a Mycobacterium
 tuberculosis antigen of the invention. The polypeptides are useful for
 immunotherapy to treat or prevent tuberculosis (especially in humans),
 e.g. they can be included with an acceptable carrier in pharmaceutical,
 compositions or included in vaccines, and administered to induce
 protective immunity in a patient against M. tuberculosis. Tuberculosis is
 a chronic, infectious disease generally caused by M. tuberculosis
 infection, and if left untreated typically results in serious
 complications and death. Fusion proteins containing the antigen, or DNA
 molecules can similarly be included with an acceptable carrier in
 pharmaceutical compositions or in vaccines and administered as above. The
 polypeptides are also useful for diagnosis of tuberculosis, by contacting
 dermal cells with at least one polypeptide and detecting an immune
 response (especially induration) on the patient's skin. Inhibiting the
 spread of tuberculosis requires vaccination and accurate diagnosis, since
 antibiotic therapy may not be effective due to the existence of an
 asymptomatic but contagious stage and to patient non-compliance. The
 polypeptides overcome concerns of safety and efficacy of current
 vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
 lack of sensitivity and specificity of existing diagnostic techniques.
 Sequence 392 AA;
 Query Match 71.7%; Score 1963; DB 20; Length 392;
 Best Local Similarity 99.7%; Pred. No. 3e-155;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 148 MSRAFIIDPTISAIDGLYDLIGIPNQGGLYSSLEYFEKALEELAAFPDGLWLSAA 207
 Db 1 MSRAFIIDPTISAIDGLYDLIGIPNQGGLYSSLEYFEKALEELAAFPDGLWLSAA 60
 QY 208 DYAGKRNHNVFFQELADLDRLQISLIHQANAVQTRDILEGAKGLEFVRPVAVDLT 267
 Db 61 DYAGKRNHNVFFQELADLDRLQISLIHQANAVQTRDILEGAKGLEFVRPVAVDLT 120
 QY 268 YIPVVGHALSAFQAPFCAGAMAVVGGALAYLVVKTINATOLLKLLAKLAELVAAATAD 327
 Db 121 YIPVVGHALSAFQAPFCAGAMAVVGGALAYLVVKTINATOLLKLLAKLAELVAAATAD 180
 QY 328 IISDVADIIGKILGEVWEFITTALNGKELMDKLTGWVTGLFSRWGNSNLESFFAGVPLGT 387
 Db 181 IISDVADIIGKILGEVWEFITTALNGKELMDKLTGWVTGLFSRWGNSNLESFFAGVPLGT 240

QY 388 GATSGLSQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGLPSLAQVHAA 447
DB 241 GATSGLSQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGLPSLAQVHAA 300
QY 448 STRQALRPADGPGVGAARQVGGQSLVSAQSGQMGPGPVGMGMHPSSGASKGTTTKY 507
DB 301 STRQALRPADGPGVGAARQVGGQSLVSAQSGQMGPGPVGMGMHPSSGASKGTTTKY 360
QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
DB 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 3

AAW73655
ID AAW73655 standard; Protein; 392 AA.

XX AC AAW73655;

XX DT 24-MAR-1999 (first entry)

XX DE M. tuberculosis antigen clone htcc#1 protein sequence.

XX KW Antigen; M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis; infection.

XX OS Mycobacterium tuberculosis.

XX PN WO9853076-A2.

XX PD 26-NOV-1998.

XX PF 20-MAY-1998; 98WO-US10514.

XX PR 05-MAY-1998; 98US-0073009.

XX PR 20-MAY-1997; 97US-0858998.

XX PA (CORI-) CORIXA CORP.

XX PI Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAW;

XX DR WPI; 1999-045315/04.

XX DR N-PSDB; AAX01143.

XX PT New isolated Mycobacterium tuberculosis antigens - used to develop products for the prevention, treatment and diagnosis of tuberculosis infection

XX PS Claim 2; Page 88-89; 104pp; English.

XX CC This sequence represents an immunogenic portion of a Mycobacterium tuberculosis antigen of the invention. The polypeptides are useful for immunotherapy to treat or prevent tuberculosis (especially in humans), e.g. they can be included with an acceptable carrier in pharmaceutical compositions or included in vaccines, and administered to induce protective immunity in a patient against M. tuberculosis. Tuberculosis is a chronic, infectious disease generally caused by M. tuberculosis infection, and if left untreated typically results in serious complications and death. Fusion proteins containing the antigen, or DNA molecules can similarly be included with an acceptable carrier in pharmaceutical compositions or in vaccines and administered as above. The polypeptides are also useful for diagnosis of tuberculosis, by contacting dermal cells with at least one polypeptide and detecting an immune response (especially induration) on the patient's skin. Inhibiting the spread of tuberculosis requires vaccination and accurate diagnosis, since antibiotic therapy may not be effective due to the existence of an asymptomatic but contagious stage and to patient non-compliance. The polypeptides overcome concerns of safety and efficacy of current vaccination with live bacteria (usually Bacillus Calmette-Guerin) and lack of sensitivity and specificity of existing diagnostic techniques.

XX SQ Sequence 392 AA;

Query Match

Best Local Similarity 71.7%; Score 1963; DB 20; Length 392;

Matches 391; Conservative 99.7%; Pred. No. 3e-155;

Mismatches 0; Indels 1; Gaps 0;

QY 148 MSRAPIIPTISAIDGLYDLIGIGIPNQGGLYSSLEYFEKALEBELAAAFPGDGLGSA 207
DB 1 MSRAPIIPTISAIDGLYDLIGIGIPNQGGLYSSLEYFEKALEBELAAAFPGDGLGSA 60
QY 208 DKYAGKRNHNFFOELADLDRLISLHDOANAVQTTTRDILEGAKKGLFVRPVAVDLT 267
DB 61 DKYAGKRNHNFFOELADLDRLISLHDOANAVQTTTRDILEGAKKGLFVRPVAVDLT 120
QY 268 YIPVVGHALSAAFPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAKLAELVAAAIAD 327
DB 121 YIPVVGHALSAAFPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAKLAELVAAAIAD 180
QY 328 IISDVADIIKGLGEVWEFITNALNGLKELMDKLTGWVTGLFSRGSNLESFFAGVPGLT 387
DB 181 IISDVADIIKGLGEVWEFITNALNGLKELMDKLTGWVTGLFSRGSNLESFFAGVPGLT 240
QY 388 GATSGLSQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGLPSLAQVHAA 447
DB 241 GATSGLSQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGLPSLAQVHAA 300
QY 448 STRQALRPADGPGVGAARQVGGQSLVSAQSGQMGPGPVGMGMHPSSGASKGTTTKY 507
DB 301 STRQALRPADGPGVGAARQVGGQSLVSAQSGQMGPGPVGMGMHPSSGASKGTTTKY 360
QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
DB 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 4

AAU08226

ID AAU08226 standard; Protein; 392 AA.

XX AC AAU08226;

XX DT 17-DEC-2001 (first entry)

XX DE Polypeptide encoded by Mycobacterium tuberculosis clone htcc#1.

XX KW Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial; immunostimulant; clone htcc.

XX OS Mycobacterium tuberculosis.

XX PN WO200162893-A2.

XX PD 30-AUG-2001.

XX PF 26-FEB-2001; 2001WO-US05992.

XX PR 25-FEB-2000; 2000US-0185037.

XX PR 08-AUG-2000; 2000US-0223828.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-Neto A, Skeiky Y, Ovendaie P, Jen S, Lodes M;

XX DR WPI; 2001-536638/59.

XX DR N-PSDB; AAS12487.

XX PT An isolated polypeptide comprising a Mycobacterium antigen, e.g., from Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune response to and inhibiting development of a Mycobacterium infection - Example 1; Page 156-157; 161pp; English.

XX CC The present invention relates to the isolation of Mycobacterium tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids

CC encoding them. The invention describes compounds and methods for the
 CC diagnosis of tuberculosis or for inducing protective immunity against
 CC tuberculosis. The compounds comprise at least one immunogenic portion
 CC of one or more Mycobacterium proteins and nucleic acid molecules
 CC encoding such polypeptides. The Mycobacterium proteins and nucleic acid
 CC molecules encoding them can be used in diagnostic kits for the detection
 CC of Mycobacterium infection in patients and biological samples. The
 CC compounds of the invention and antibodies directed against the
 CC Mycobacterium proteins may be used in vaccines for immunisation against
 CC Mycobacterium infections. The nucleic acids encoding the Mycobacterium
 CC proteins may be used in gene therapy. The present sequence represents
 CC the polypeptide encoded by M. tuberculosis clone H7C#1.
 XX
 SQ Sequence 392 AA;

Query Match 71.7%; Score 1963; DB 22; Length 392;
 Best Local Similarity 99.7%; Pred. No. 3e-155;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALELAAFPDGLGSA 207
 Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALELAAFPDGLGSA 60

QY 208 DKYAGNRNHNFFQELADLRQLSLIHQANAVQTRDILEGAKGLEFVRPVA 267
 Db 61 DKYAGNRNHNFFQELADLRQLSLIHQANAVQTRDILEGAKGLEFVRPVA 120

QY 268 YIPVVGHALSAFAFPFCAGAMAVVGGALVYLVKTLINATQLKLAELVAA 327
 Db 121 YIPVVGHALSAFAFPFCAGAMAVVGGALVYLVKTLINATQLKLAELVAA 180

QY 328 IISDVADIIGKLGVEWFFITNALNGLKELMDKLTGVTGLFSRGWSNLEFF 387
 Db 181 IISDVADIIGKLGVEWFFITNALNGLKELMDKLTGVTGLFSRGWSNLEFF 240

QY 388 GATSGLSQVTGTFGAAGLSASSGLAHADSLASSALPALAGIGGSGFGG 447
 Db 241 GATSGLSQVTGTFGAAGLSASSGLAHADSLASSALPALAGIGGSGFGG 300

QY 448 STRQALRPADGPGVGAAGAEVGGQSQLVSAQSGQGVGMGMHPSSGASKG 507
 Db 301 STRQALRPADGPGVGAAGAEVGGQSQLVSAQSGQGVGMGMHPSSGASKG 360

QY 508 SEGAACTEDARAPVEADAGGQKVLVRNV 539
 Db 361 SEGAACTEDARAPVEADAGGQKVLVRNV 392

RESULT 5
 AAU01882
 ID AAU01882 standard; Protein; 392 AA.
 XX
 AC AAU01882;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE M. tuberculosis antigen HTCC#1/Mtb40.
 XX
 KW HTCC#1; Mtb40; antigen; vaccine; tuberculosis; AIDS;
 XX acquired immunodeficiency disease.
 OS Mycobacterium tuberculosis.
 FH Key Location/Qualifiers
 FT Peptide 1..20
 FT /label= Peptide 1
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 14..33
 FT /label= Peptide 2
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 27..46
 FT /label= Peptide 3
 FT /note= "Used for T-cell epitope mapping"

FT Peptide 40..59
 FT /label= Peptide 4
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 53..72
 FT /label= Peptide 5
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 66..85
 FT /label= Peptide 6
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 79..98
 FT /label= Peptide 7
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 92..111
 FT /label= Peptide 8
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 105..124
 FT /label= Peptide 9
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 118..137
 FT /label= Peptide 10
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 131..150
 FT /label= Peptide 11
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 144..163
 FT /label= Peptide 12
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 157..176
 FT /label= Peptide 13
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 170..189
 FT /label= Peptide 14
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 183..202
 FT /label= Peptide 15
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 196..215
 FT /label= Peptide 16
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 209..228
 FT /label= Peptide 17
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 222..241
 FT /label= Peptide 18
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 235..254
 FT /label= Peptide 19
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 248..267
 FT /label= Peptide 20
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 261..280
 FT /label= Peptide 21
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 274..293
 FT /label= Peptide 22
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 287..306
 FT /label= Peptide 23
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 300..319
 FT /label= Peptide 24
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 313..332
 FT /label= Peptide 25
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 325..344
 FT /label= Peptide 26
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 339..358
 FT /label= Peptide 27
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 352..371

```

FT /label= Peptide_28
FT /note= "Used for T-cell epitope mapping"
FT Peptide 365..384
FT /label= Peptide_29
FT /note= "Used for T-cell epitope mapping"
FT Peptide 373..392
FT /label= Peptide_30
FT /note= "Used for T-cell epitope mapping"
FT XX
FT WO200124820-A1.
FN
XX
XX 12-APR-2001.
XX
XX 10-OCT-2000; 2000WO-US28095..
XX
XX 07-OCT-1999; 99US-0158338.
PR 07-OCT-1999; 99US-0158338.
XX 07-OCT-1999; 99US-0158425.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX WPI; 2001-290576/30.
DR N-PSDB; AAS03773.
DR
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
XX proteins comprising combinations of heterologous antigens -
XX
XX Example 2; Fig 6; 168pp; English.
XX
XX The sequence represents Mycobacterium tuberculosis H7CC#1 (also known
XX as Mtb40), an M. tuberculosis antigen. Compositions comprising at least
XX 2 heterologous antigens, as a fusion protein, and vectors expressing the
XX fusion proteins are used as vaccines to prophylactically immunise
XX mammals (especially humans) against infection by Mycobacteria. The
XX compositions contain at least 2 heterologous antigens that increase the
XX serological sensitivity of individuals infected with tuberculosis, a
XX disease frequently affecting patients with acquired immunodeficiency
XX disease, AIDS.
XX
XX Sequence 392 AA;
SQ
Query Match 71.7%; Score 1963; DB 22; Length 392;
Best Local Similarity 99.7%; Pred. No. 3e-155;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 148 MSRAFIIDPTTSAIDGLYDLIGIGIPNOGGILYSSLEYFEKALELAAAFPGDGLGSA 207
Db 1 MSRAFIIDPTTSAIDGLYDLIGIGIPNOGGILYSSLEYFEKALELAAAFPGDGLGSA 60
QY 208 DKYAGKRNHNVNFFQELADLRQLISLIHQDQNAVQTTDRILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNVNFFQELADLRQLISLIHQDQNAVQTTDRILEGAKKGLFVRPVAVDLT 120
QY 268 YIPVVGHALSAAFPQPCAGAMVVGGLAYLVVKTLLINATQLLKLAKLAEVAAAIAD 327
Db 121 YIPVVGHALSAAFPQPCAGAMVVGGLAYLVVKTLLINATQLLKLAKLAEVAAAIAD 180
QY 328 IISDVADIITKGLGEVWEFITTALNGKLWDKLTGWVTGLFSRGNLSFFRAGVPGLT 387
Db 191 IISDVADIITKGLGEVWEFITTALNGKLWDKLTGWVTGLFSRGNLSFFRAGVPGLT 240
QY 388 GATSGLSQVTVGLFGAAGLSASGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 241 GATSGLSQVTVGLFGAAGLSASGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
QY 448 STQRLRPRADGPVGAARQVGGOSQVLSAOGSGMGPGVCMGGMHPSSGASKGTTTKKY 507
Db 301 STQRLRPRADGPVGAARQVGGOSQVLSAOGSGMGPGVCMGGMHPSSGASKGTTTKKY 360
QY 508 SEGAAAGTDEARAPVEADAGGQKVLVRNVV 539
Db 361 SEGAAAGTDEARAPVEADAGGQKVLVRNVV 392

```

RESULT 6

ABU05959
ID ABU05959 standard; Protein; 392 AA.

XX AC ABU05959;

XX DT 08-APR-2003 (first entry)

XX DE M. tuberculosis and M. leprae marker protein #610.

XX KW Mycobacterioses; survival; virulence; protective antigen; vaccine;
KW mycobacterial disease; tuberculosis; leprosy.

XX OS Mycobacterium tuberculosis.

XX OS Mycobacterium leprae.

XX EN WO200274903-A2.

XX PD 26-SEP-2002.

XX PF 22-FEB-2002; 2002WO-IB01973.

XX PR 22-FEB-2001; 2001US-270123P.

XX PA (INSP) INST PASTEUR.

XX PI Cole S;

XX WPI; 2002-759885/82.

XX PT Identifying and selecting genes for survival or virulence of
PT mycobacteria by a comparative genomic analysis of the sequences of
PT Mycobacterium tuberculosis and M. leprae -

XX PS Claim 17; Page 828; 874pp; English.

CC This invention relates to a novel method for identifying essential genes
CC for survival or virulence of mycobacteria species. The method comprises
CC aligning the genomic sequence of a first mycobacterium species on a
CC genomic sequence of a second mycobacterium species and selecting a
CC polynucleotide sequence that is highly conserved in both genomes with no
CC counterparts in other bacterial genomic sequences and that corresponds
CC to an essential gene for the survival or virulence of mycobacterium
CC species. The method of the invention is useful for detecting M.
CC tuberculosis or M. leprae infection. The method reduces the number of
CC potential new targets and protective antigens for new drugs and vaccine
CC compositions to treat and prevent mycobacterial diseases, particularly
CC tuberculosis and leprosy. The present sequence represents a marker
CC protein from Mycobacterium tuberculosis and Mycobacterium leprae
CC identified using the method of the invention.

XX SQ Sequence 392 AA;

Query Match 71.7%; Score 1963; DB 23; Length 392;
Best Local Similarity 99.7%; Pred. No. 3e-155;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTTSAIDGLYDLIGIGIPNOGGILYSSLEYFEKALELAAAFPGDGLGSA 207

Db 1 MSRAFIIDPTTSAIDGLYDLIGIGIPNOGGILYSSLEYFEKALELAAAFPGDGLGSA 60

QY 208 DKYAGKRNHNVNFFQELADLRQLISLIHQDQNAVQTTDRILEGAKKGLFVRPVAVDLT 267

Db 61 DKYAGKRNHNVNFFQELADLRQLISLIHQDQNAVQTTDRILEGAKKGLFVRPVAVDLT 120

QY 268 YIPVVGHALSAAFPQPCAGAMVVGGLAYLVVKTLLINATQLLKLAKLAEVAAAIAD 327

Db 121 YIPVVGHALSAAFPQPCAGAMVVGGLAYLVVKTLLINATQLLKLAKLAEVAAAIAD 180

QY 328 IISDVADIITKGLGEVWEFITTALNGKLWDKLTGWVTGLFSRGNLSFFRAGVPGLT 387

Db 181 IISDVADIKGTGGEVWEFTNALNGKELWDLKLTGWVTGLFSRGSNLESFFAGVPGLT 240
 QY 388 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
 Db 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 QY 448 STROALRPRADGPVGAAGAAEQVGGQSLVSAQSQGSGGVGNGHPSGASKGTTTKKY 507
 Db 301 STROALRPRADGPVGAAGAAEQVGGQSLVSAQSQGSGGVGNGHPSGASKGTTTKKY 360
 QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 539
 Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 7
 AAE29714
 ID AAE29714 standard; Protein; 392 AA.
 XX
 AC AAE29714;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium sp. HTCC#1 antigenic protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; HTCC#1 antigen; MTB40.
 XX
 OS Mycobacterium sp.
 XX
 PN WO200272792-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 13-MAR-2002; 2002WO-US08223.
 XX
 PR 13-MAR-2001; 2001US-275837P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Brannon M, Guderian J;
 XX
 DR WPI; 2002-759844/82.
 DR N-PSDB; AAD47091.
 XX
 PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis
 XX
 PS Disclosure; Page 103-104; 155pp; English.
 XX
 SS The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention are used in methods for eliciting immune response in mammals. They are useful as vaccines to elicit protective immunity against pathogenic microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion polypeptides are used for enhancing the expression of polynucleotides, as in vivo diagnostic agents and for raising antibodies in a non-human animal. The invention is used in gene therapy. The present sequence is Mycobacterium sp. HTCC#1 antigenic protein. HTCC#1 is also referred to as MTB40.
 XX
 SQ Sequence 392 AA;

Query Match 71.7%; Score 1963; DB 23; Length 392;
 Best Local Similarity 99.7%; Pred. No. 3e-155;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTISALDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAPPGDWLGSAA 207
 Db 1 MSRAFIIDPTISALDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAPPGDWLGSAA 60
 QY 208 DKVAGKRNHNFFQELADLDROLISLIHDQANAVOTTRDILLEGAKKGLFVRPVAVDLT 267
 Db 61 DKVAGKRNHNFFQELADLDROLISLIHDQANAVOTTRDILLEGAKKGLFVRPVAVDLT 120
 QY 268 YIPVGHASAAFPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAELVAAAIAD 327
 Db 121 YIPVGHASAAFPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAELVAAAIAD 180
 QY 328 IISDVADIKGTGGEVWEFTNALNGKELWDLKLTGWVTGLFSRGSNLESFFAGVPGLT 387
 Db 181 IISDVADIKGTGGEVWEFTNALNGKELWDLKLTGWVTGLFSRGSNLESFFAGVPGLT 240
 QY 388 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
 Db 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 QY 448 STROALRPRADGPVGAAGAAEQVGGQSLVSAQSQGSGGVGNGHPSGASKGTTTKKY 507
 Db 301 STROALRPRADGPVGAAGAAEQVGGQSLVSAQSQGSGGVGNGHPSGASKGTTTKKY 360
 QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 539
 Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 8
 AAE17578
 ID AAE17578 standard; Protein; 392 AA.
 XX
 AC AAE17578;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium species MTB40 (HTCC #1) protein.
 XX
 KW Fusion protein; antigen; serological sensitivity; immune response; tuberculosis; infection; vaccine; MTB40; HTCC #1 protein.
 XX
 OS Mycobacterium sp.
 XX
 PN WO200198460-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 20-JUN-2001; 2001WO-US19959.
 XX
 PR 20-JUN-2000; 2000US-0597796.
 PR 01-FEB-2001; 2001US-265737P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Alderson M;
 XX
 DR WPI; 2002-147798/19.
 DR N-PSDB; AAD28349.
 XX
 PT Composition comprising MTB39 antigen and MTB32A antigen from Mycobacterium species, useful for eliciting immune response in a subject
 XX
 PS Claim 9; Page 120; 136pp; English.
 XX
 SS The present invention relates to fusion proteins containing at least two Mycobacterium species antigens, nucleotides encoding them and compositions comprising such fusion proteins. The present invention particularly relates to nucleic acids encoding fusion proteins that include two or more individual M. tuberculosis antigens which increase the serological sensitivity of sera from individuals infected with tuberculosis and methods for their use in diagnosis, prevention and

CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected
 CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB40 (HTCC #1) protein.

XX SQ Sequence 392 AA;

Query Match 71.7%; Score 1963; DB 22; Length 392;
 Best Local Similarity 99.7%; Pred. No. 3e-155;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 148 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEBELAAFPDGLGSA 207
 DB 1 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEBELAAFPDGLGSA 60
 QY 208 DKYAGKRNHNVNFFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 267
 DB 61 DKYAGKRNHNVNFFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 120
 QY 268 YIPVVGHALSAFAQPFACAGAMVVGALAYLVVKTLLNATQLLKLAELVAAD 327
 DB 121 YIPVVGHALSAFAQPFACAGAMVVGALAYLVVKTLLNATQLLKLAELVAAD 180
 QY 328 IISVDVADIIKGLGEVWEFITNALNGLKELWDLTGHWVTGLFSRGWSNLESFFAGVPGLT 387
 DB 181 IISVDVADIIKGLGEVWEFITNALNGLKELWDLTGHWVTGLFSRGWSNLESFFAGVPGLT 240
 QY 388 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
 DB 241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 QY 448 STROALPRADPGPVGAAGAEQVGGOSQLVSAQSGQMGPGVGMGMHPSSGASKGTTTKY 507
 DB 301 STROALPRADPGPVGAAGAEQVGGOSQLVSAQSGQMGPGVGMGMHPSSGASKGTTTKY 360
 QY 508 SEGAAAGTDAERAPVEADAGGOKVLVRNVV 539
 DB 361 SEGAAAGTDAERAPVEADAGGOKVLVRNVV 392

RESULT 9

ID AAU01886
 XX AAU01886 standard; Protein; 408 AA.
 AC AAU01886;
 XX 29-AUG-2001 (first entry)
 DT M. tuberculosis antigen HTCC#1(TM-1).
 XX HTCC#1(TM-1); antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease; His Tag.
 XX Mycobacterium tuberculosis.
 OS
 XX Key Location/Qualifiers
 FH Binding-site 1..8
 FT /label= Histidine tag
 FT /note= "Nickel chelating region used to aid
 FT purification of the protein"
 FT Misc-difference 169..180
 FT /note= "These residues are deleted to form the
 FT HTCC#1(TM-1) protein sequence"

FT Misc-difference 405
 FT /label= OTHER
 FT /note= "In frame STOP codon"

XX WO200124820-A1.

XX 12-APR-2001.

PD 10-OCT-2000; 2000WO-US28095.

PF 07-OCT-1999; 99US-0158338.

PR 07-OCT-1999; 99US-0158425.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;

XX WPI; 2001-290576/30.

DR N-PSDB; AAS03777.

XX Vaccinating against Mycobacteria infections in mammals using fusion
 FT proteins comprising combinations of heterologous antigens -

XX Example 2; Fig 9a; 168pp; English.

XX The sequence is Mycobacterium tuberculosis HTCC#1(TM-1)
 CC representing H7C#1 (an M. tuberculosis antigen) lacking the first
 CC transmembrane domain and includes a His tag at the N-terminus to aid
 CC purification. Compositions comprising at least 2 heterologous antigens,
 CC as a fusion protein, and vectors expressing the fusion proteins are used
 CC as vaccines to prophylactically immunise mammals (especially humans)
 CC against infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS.

XX SQ Sequence 408 AA;

Query Match 70.9%; Score 1941; DB 22; Length 408;
 Best Local Similarity 98.7%; Pred. No. 2.1e-153;
 Matches 391; Conservative 0; Mismatches 1; Indels 4; Gaps 2;

QY 148 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEBELAAFPDGLGSA 207
 DB 9 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEBELAAFPDGLGSA 68
 QY 208 DKYAGKRNHNVNFFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 267
 DB 69 DKYAGKRNHNVNFFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 128
 QY 268 YIPVVGHALSAFAQPFACAGAMVVGGA--LAYLVVKTLLNA--TOLLKLAELVA 323
 DB 129 YIPVVGHALSAFAQPFACAGAMVVGGA--LAYLVVKTLLNA--TOLLKLAELVA 188
 QY 324 AIADIISDVADIIKGLGEVWEFITNALNGLKELWDLTGHWVTGLFSRGWSNLESFFAGV 383
 DB 189 AIADIISDVADIIKGLGEVWEFITNALNGLKELWDLTGHWVTGLFSRGWSNLESFFAGV 248
 QY 384 PGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 443
 DB 249 PGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 308
 QY 444 VHAASSTROALPRADPGPVGAAGAEQVGGOSQLVSAQSGQMGPGVGMGMHPSSGASKGTT 503
 DB 309 VHAASSTROALPRADPGPVGAAGAEQVGGOSQLVSAQSGQMGPGVGMGMHPSSGASKGTT 368
 QY 504 TKKYSAGAAAGTDAERAPVEADAGGOKVLVRNVV 539
 DB 369 TKKYSAGAAAGTDAERAPVEADAGGOKVLVRNVV 404

RESULT 10
 AAU01887

ID AAU01887 standard; Protein; 408 AA.
 AC AAU01887;
 DT 29-AUG-2001 (first entry)
 DE M. tuberculosis antigen HTCC#1(TM-2).
 KW HTCC#1(TM-2); antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease; His Tag.
 OS Mycobacterium tuberculosis.
 PH Key
 FT Binding-site 1..8 Location/Qualifiers
 FT /label= Histidine tag
 FT /note= "Nickel chelating region used to aid
 FT purification of the protein"
 FT Misc-difference 110..213
 FT /note= "These residues are deleted to form the
 FT HTCC#1(TM-2) protein sequence"
 FT Misc-difference 405
 FT /label= OTHER
 FT /note= "In frame STOP codon"
 XX
 PN WO200124820-A1.
 XX
 PD 12-APR-2001.
 XX
 PF 10-OCT-2000; 2000WO-US28095.
 XX
 PR 07-OCT-1999; 99US-0158338.
 PR 07-OCT-1999; 99US-0158425.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX WPI; 2001-290576/30.
 DR N-PSDB; AAS03778.
 XX
 PT Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 XX
 PS Example 2; Fig 9d; 168pp; English.
 XX
 CC The sequence is Mycobacterium tuberculosis HTCC#1(TM-1)
 CC representing HTCC#1 (an M. tuberculosis antigen) lacking all the
 CC transmembrane domains and includes a His tag at the N-terminus to aid
 CC purification. Compositions comprising at least 2 heterologous antigens,
 CC as a fusion protein, and vectors expressing the fusion proteins are used
 CC as vaccines to prophylactically immunise mammals (especially humans)
 CC against infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS.
 XX
 SQ Sequence 408 AA;
 Query Match 70.9%; Score 1941; DB 22; Length 408;
 Best Local Similarity 98.7%; Pred. No. 2.1e-153;
 Matches 391; Conservative 0; Mismatches 1; Indels 4; Gaps 2;
 QY 148 MSRAFIIDPTISAIDGLYDILLGIPNCGGILYSSLEYFEKALEBLAAFFGDGWLGSAA 207
 DB 9 MSRAFIIDPTISAIDGLYDILLGIPNCGGILYSSLEYFEKALEBLAAFFGDGWLGSAA 68
 QY 208 DKYAGKRNHNVFFQSLADRLQLSLIHDOANVCTTRD--ILEGAKKGLFEFVRPVAVD 265
 DB 69 DKYAGKRNHNVFFQSLADRLQLSLIHDOANVCTTRDKLILEGAKKGLFEFVRPVAVD 128
 QY 266 LTYIPVVGHALSAAFOAPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAKLAELVAAAI 325

DB 129 LTYIPVVGHALSAAFOAPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAKLAELVAAAI 188
 QY 326 ADIISDVADIIGKILGEVWEFITNA--LNGKLKELMDKLTGWVTGLFSRGWSNLESFFAGV 383
 DB 189 ADIISDVADIIGKILGEVWEFITNAKLNLGLKELMDKLTGWVTGLFSRGWSNLESFFAGV 248
 QY 384 PGLTGATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 443
 DB 249 PGLTGATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 308
 QY 444 VHAASTRQALRPRAADGVGAAAEQVGGQSQLVSAQSQGMGFVGMGMHPPSSGASKGTT 503
 DB 309 VHAASTRQALRPRAADGVGAAAEQVGGQSQLVSAQSQGMGFVGMGMHPPSSGASKGTT 368
 QY 504 TKKYSAGAAAGTHEDAEAPVEADAGGGQKVLVRNVV 539
 DB 369 TKKYSAGAAAGTHEDAEAPVEADAGGGQKVLVRNVV 404
 RESULT 11
 AAU01903
 ID AAU01903 standard; Protein; 788 AA.
 AC AAU01903;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE M. tuberculosis antigen HTCC#1 fusion protein #2.
 XX
 KW HTCC#1(1-149)-TbH9-HTCC#1(161-392); antigen; vaccine; tuberculosis;
 KW AIDS; acquired immunodeficiency disease; His Tag.
 XX
 OS Mycobacterium tuberculosis.
 OS Synthetic.
 XX
 PH Key
 FT Binding-site 3..8 Location/Qualifiers
 FT /label= Histidine tag
 FT /note= "Nickel chelating region used to aid
 FT purification of the protein"
 FT Misc-difference 785
 FT /label= OTHER
 FT /note= "In frame STOP codon"
 XX
 PN WO200124820-A1.
 XX
 PD 12-APR-2001.
 XX
 PF 10-OCT-2000; 2000WO-US28095.
 XX
 PR 07-OCT-1999; 99US-0158338.
 PR 07-OCT-1999; 99US-0158425.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX WPI; 2001-290576/30.
 DR Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 XX
 PS Example 2; Fig 11; 168pp; English.
 XX
 CC The sequence represents Mycobacterium tuberculosis fusion protein,
 CC HTCC#1(1-149)-TbH9-HTCC#1(161-392) and includes a His tag at the
 CC N-terminus to aid purification. Compositions comprising at least 2
 CC heterologous antigens, as a fusion protein, and vectors expressing the
 CC fusion proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease.

```
CC AIDS.
XX
SQ Sequence 788 AA;
Query Match 62.8%; Score 1718; DB 22; Length 788;
Best Local Similarity 49.5%; Pred. No. 2.2e-134;
Matches 384; Conservative 2; Mismatches 6; Indels 384; Gaps 2;
QY 148 MSRAFIIDPTISAIDGLYDLIGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
DB 9 MSRAFIIDPTISAIDGLYDLIGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 68
QY 208 DXYAGKRNHNVFFQELADLRQLISLIHDOANAVQTTDRILEGAKKGLFVRPVAVDLT 267
DB 69 DXYAGKRNHNVFFQELADLRQLISLIHDOANAVQTTDRILEGAKKGLFVRPVAVDLT 128
QY 268 YIPVVGHALSAAFQAPFCAGAMVVGALAYLV 300
DB 129 YIPVVGHALSAAFQAPFCAGAMVVGALAYLV 188
QY 301 300
DB 189 QMDSVADLPSAASFQSVVWGLTVGSSAGLMVAAPVYVAMVSVTAGQAELETA 248
QY 301 300
DB 249 QVRVAABAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPAIVNEAYEGEMWAQDA 308
QY 301 305
DB 309 AAMFGYAAATATATATLIPFEAPEMTSAGLLQEAAAEEASPTAAANGLMNNVPOALQ 368
QY 306 305
DB 369 QLAQPTQTTSSKLGGLKWTVSPHRSPISNNWGMANNHSMNTGVSMTNTLSMLKGF 428
QY 306 305
DB 429 AAAAAQAQVTAQAQNGVRAMSSGLSSGLGGVAAANLGRAASVGLSVPOQAAANQ 488
QY 306 305
DB 489 AVTFAARALPITSAAERGGQVGLGFLPVQWGARAGGLGVLVPRPPRYVPHSPA 548
QY 306 363
DB 549 AGKLTQLLKLAKLAELVAAAIADIISDVADIINKGILGEVWEFTINALNGKELWDKLTG 608
QY 364 WVTGLFSRGWNLSEFPAGVPLGTCATSGLSQVTLGFCAGLSASSGLAHADSLASS 423
DB 609 WVTGLFSRGWNLSEFPAGVPLGTCATSGLSQVTLGFCAGLSASSGLAHADSLASS 668
QY 424 PALAGIGGGSGFGLPSLAQVHAASSTQALRPRADGPVGAAGAAEOVGSGOLVSAQSGQM 483
DB 669 PALAGIGGGSGFGLPSLAQVHAASSTQALRPRADGPVGAAGAAEOVGSGOLVSAQSGQM 728
QY 484 GGPVGMGMGHPSSGASKGTTTTKKYSEGAAGCTEDAERAPVEADAGGQKVLVRNV 539
DB 729 GGPVGMGMGHPSSGASKGTTTTKKYSEGAAGCTEDAERAPVEADAGGQKVLVRNV 784
RESULT 12
ABU05402
ID ABU05402 standard; Protein; 394 AA.
XX
AC ABU05402;
XX
DT 08-APR-2003 (first entry)
XX
DE M. tuberculosis and M. leprae marker protein #53.
XX
KW Mycobacterioses; survival; virulence; protective antigen; vaccine;
XX mycobacterial disease; tuberculosis; leprosy.
```

```
XX Mycobacterium tuberculosis.
OS Mycobacterium leprae.
XX
PN WO200274903-A2.
XX
PD 26-SEP-2002.
XX
XX 22-FEB-2002; 2002WO-IB01973.
XX
XX 22-FEB-2001; 2001US-270123P.
XX
PA (INSP ) INST PASTEUR.
XX
PI Cole S;
XX
DR WPI; 2002-759885/82.
XX
PT Identifying and selecting genes for survival or virulence of
PT mycobacteria by a comparative genomic analysis of the sequences of
PT Mycobacterium tuberculosis and M. leprae -
XX
PS Claim 17; Page 224-225; 874pp; English.
XX
CC This invention relates to a novel method for identifying essential genes
CC for survival or virulence of mycobacteria species. The method comprises
CC aligning the genomic sequence of a first mycobacterium species on a
CC genomic sequence of a second mycobacterium species and selecting a
CC polynucleotide sequence that is highly conserved in both genomes with no
CC counterparts in other bacterial genomic sequences and that corresponds
CC to an essential gene for the survival or virulence of mycobacterium
CC species. The method of the invention is useful for detecting M.
CC tuberculosis or M. leprae infection. The method reduces the number of
CC potential new targets and protective antigens for new drugs and vaccine
CC compositions to treat and prevent mycobacterial diseases, particularly
CC tuberculosis and leprosy. The present sequence represents a marker
CC protein from Mycobacterium tuberculosis and Mycobacterium leprae
CC identified using the method of the invention.
XX
SQ Sequence 394 AA;
Query Match 46.1%; Score 1263; DB 23; Length 394;
Best Local Similarity 62.9%; Pred. No. 7.8e-97;
Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;
QY 148 MSRAFIIDPTISAIDGLYDLIGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
DB 1 MCGAFIIDPTLKAIEAWHALLGIVNDGCVLYSSLFEEKALEHAAAFPGDGLGSA 60
QY 208 DXYAGKRNHNVFFQELADLRQLISLIHDOANAVQTTDRILEGAKKGLFVRPVAVDLT 267
DB 61 DXYAGKRNKRVDFQELADLRQLISLIHDOANAVQTTDRILEGAKKGLFVRPVAIDLN 120
QY 268 YIPVVGHALSAAFQAPFCAGAMVVGALAYLVVKTLLINATOLLKLAKLAELVAAAIAD 327
DB 121 YIPVVGSMVMSASTQACAAAVVSGGLAYLLVQTAHTAKFVALLARLHALLASAVAD 180
QY 328 IISDVADIINKGILGEVWEFTINALNGKELWDKLTGWTGTLFSRGWNLSEFPAGVPLGT 387
DB 181 VVSDGVAILIKGIVDHLWHFIAGLTGLKDIVKIIHWFGFLFSHWNLSHLSFFGGLPGLS 240
QY 388 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSSLPALAGIGGGSGFGLPSLAQVHAA 447
DB 241 GATSGLSQVTLGFGVPGLAGSSGLSSTENLSTENLSTENLSTENLSTENLSTENLSTEN 300
QY 448 STQALRPRADGPVGAAGAAEOVGSGOLVSAQSGQVGVGMGMGHPSSGASK - GTTTK 505
DB 301 STQGTTRSQGVSAELSTEQFGQGEFVSAQSGQVGVGMGMGHPSSGASK - GTTTK 360
QY 506 KYSEGAAGCTEDAERAPVEADAGGQKVLVRNV 539
DB 361 KYSEGAAGCTEDAERAPVEADAGGQKVLVRNV 394
```

RESULT 13
 AAU01883
 ID AAU01883 standard; Protein; 242 AA.
 AC AAU01883;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE M. tuberculosis antigen HTCC#1 (1-223).
 XX
 KW HTCC#1 (1-223); antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease; His Tag.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Binding-site 2..7 /label= Histidine tag
 FT /note= "Nickel chelating region used to aid
 FT purification of the protein"
 FT Misc-difference 240
 FT /label= OTHER
 FT /note= "In frame STOP codon"
 FT
 PN WO200124820-A1.
 PD 12-APR-2001.
 XX
 XX 10-OCT-2000; 2000WO-US28095.
 XX
 XX 07-OCT-1999; 99US-0158338.
 PR 07-OCT-1999; 99US-0158425.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 PI WPI; 2001-290576/30.
 DR N-PSDB; AAS03774.
 DR
 XX
 PT Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 XX
 PS Example 2; Fig 7a; 168pp; English.
 CC
 CC The sequence is Mycobacterium tuberculosis HTCC#1 (1-223).
 CC representing the first transmembrane domain of HTCC#1 (an M.
 CC tuberculosis antigen) and includes a His tag at the N-terminus to aid
 CC purification. Compositions comprising at least 2 heterologous antigens,
 CC as a fusion protein, and vectors expressing the fusion proteins are used
 CC as vaccines to prophylactically immunise mammals (especially humans)
 CC against infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS.
 XX
 SQ Sequence 242 AA;
 Query Match 42.7%; Score 1168; DB 22; Length 242;
 Best Local Similarity 99.6%; Pred. No. 3 3e-89;
 Matches 231; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 148 MSRAFIIDPTISAIIDGLYDILLGIPNOGGIYSSLEYFEKALELEAAFPDGLGSA 207
 DB 8 MSRAFIIDPTISAIIDGLYDILLGIPNOGGIYSSLEYFEKALELEAAFPDGLGSA 67
 QY 208 DKYAGKRNHNVFFQELADRLQLSLIHQDQNAVQTRDILEGAKKGLFVRPVAVDLT 267
 DB 68 DKYAGKRNHNVFFQELADRLQLSLIHQDQNAVQTRDILEGAKKGLFVRPVAVDLT 127
 QY 268 YIPVVGHALSAFAQAPFCAGAMVGGALAYLVVKTILNATQLLKLAELVAAAIAD 327
 DB 9 DVADIINGILGEVWEFITNALNGKLKELWDKLTGWTGLFSRGWSNLESPFAGVGLTGAT 68

Db 128 YIPVVGHALSAFAQAPFCAGAMVGGALAYLVVKTILNATQLLKLAELVAAAIAD 187
 QY 328 IISDVADIINGILGEVWEFITNALNGKLKELWDKLTGWTGLFSRGWSNLESP 379
 XX
 Db 188 IISDVADIINGILGEVWEFITNALNGKLKELWDKLTGWTGLFSRGWSNLESP 239
 XX
 RESULT 14
 AAU01902
 ID AAU01902 standard; Protein; 744 AA.
 AC AAU01902;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE M. tuberculosis antigen HTCC#1 fusion protein #1.
 XX
 KW HTCC#1 (184-392)-TBH9-HTCC#1 (1-129); antigen; vaccine; tuberculosis;
 KW AIDS; acquired immunodeficiency disease; His Tag.
 XX
 OS Mycobacterium tuberculosis.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Binding-site 3..8 /label= Histidine tag
 FT /note= "Nickel chelating region used to aid
 FT purification of the protein"
 FT Misc-difference 742
 FT /label= OTHER
 FT /note= "In frame STOP codon"
 FT
 PN WO200124820-A1.
 PD 12-APR-2001.
 XX
 XX 10-OCT-2000; 2000WO-US28095.
 XX
 XX 07-OCT-1999; 99US-0158338.
 PR 07-OCT-1999; 99US-0158425.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 PI WPI; 2001-290576/30.
 DR N-PSDB; AAS03790.
 DR
 XX
 PT Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 XX
 PS Claim 14; Fig 10; 168pp; English.
 CC
 CC The sequence represents Mycobacterium tuberculosis fusion protein,
 CC HTCC#1 (184-392)-TBH9-HTCC#1 (1-129) and includes a His tag at the
 CC N-terminus to aid purification. Compositions comprising at least 2
 CC heterologous antigens, as a fusion protein, and vectors expressing the
 CC fusion proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS.
 XX
 SQ Sequence 744 AA;
 Query Match 38.7%; Score 1060; DB 22; Length 744;
 Best Local Similarity 100.0%; Pred. No. 1 7e-79;
 Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 331 DVADIINGILGEVWEFITNALNGKLKELWDKLTGWTGLFSRGWSNLESPFAGVGLTGAT 390
 DB 9 DVADIINGILGEVWEFITNALNGKLKELWDKLTGWTGLFSRGWSNLESPFAGVGLTGAT 68


```
Qy 391 SGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAASR 450
    |||||
Db 69 SGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAASR 128
    |||||
Qy 451 QALPRADGPGVGAABEQVGGSQLVSAQSGQGMGPGVGMGMPPSSGASGTTTKYSEG 510
    |||||
Db 129 QALPRADGPGVGAABEQVGGSQLVSAQSGQGMGPGVGMGMPPSSGASGTTTKYSEG 188
    |||||
Qy 511 AAAGTEDAERAPVEADAGGQKVLVRNV 539
    |||||
Db 189 AAAGTEDAERAPVEADAGGQKVLVRNV 217

RESULT 15
AAU01904
ID AAU01904 standard; Protein; 815 AA.
AC AAU01904;
XX
DT 29-AUG-2001 (first entry)
XX
DE M. tuberculosis antigen HTCC#1 fusion protein #3.
XX
KW HTCC#1(184-392)-TbH9-HTCC#1(1-200); antigen; vaccine; tuberculosis;
XX AIDS; acquired immunodeficiency disease; His tag.
XX
OS Mycobacterium tuberculosis.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Binding-site 3..8
FT /label= Histidine_tag
FT /note= "Nickel chelating region used to aid
FT purification of the protein"
FT Misc-difference 813
FT /label= OTHER
FT /note= "In frame STOP codon"
XX
FN WO200124820-A1.
XX
PD 12-APR-2001.
XX
PF 10-OCT-2000; 2000WO-US28095.
XX
PR 07-OCT-1999; 99US-0158338.
XX
PR 07-OCT-1999; 99US-0158425.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
XX WPI; 2001-290576/30.
XX
PT Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens -
XX
XX Example 2; Fig 12; 168pp; English.
XX
XX The sequence represents Mycobacterium tuberculosis fusion protein,
XX HTCC#1(184-392)-TbH9-HTCC#1(1-200) and includes a His tag at the
XX N-terminus to aid purification. Compositions comprising at least 2
XX heterologous antigens, as a fusion protein, and vectors expressing the
XX fusion proteins are used as vaccines to prophylactically immunise mammals
XX (especially humans) against infection by Mycobacteria. The compositions
XX contain at least 2 heterologous antigens that increase the serological
XX sensitivity of individuals infected with tuberculosis, a disease
XX frequently affecting patients with acquired immunodeficiency disease,
XX AIDS.
XX
SQ Sequence 815 AA;
```

Query Match 38.7%; Score 1060; DB 22; Length 815;

```
Best Local Similarity 100.0%; Pred. No. 1.9e-79;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 331 DVADIIKGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRCGWSNLESFFAGVPGLTGAT 390
    |||||
Db 9 DVADIIKGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRCGWSNLESFFAGVPGLTGAT 68
    |||||
Qy 391 SGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAASR 450
    |||||
Db 69 SGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAASR 128
    |||||
Qy 451 QALPRADGPGVGAABEQVGGSQLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKYSEG 510
    |||||
Db 129 QALPRADGPGVGAABEQVGGSQLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKYSEG 188
    |||||
Qy 511 AAAGTEDAERAPVEADAGGQKVLVRNV 539
    |||||
Db 189 AAAGTEDAERAPVEADAGGQKVLVRNV 217

Search completed: November 21, 2003, 16:03:21
Job time : 33.6333 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 11.3667 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-64

Perfect score: 2737

Sequence: 1 MHHHHHTAASDNFQLSQGG.....RAPVEADAGGQKVLRYNVV 539

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCFUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	1963	71.7	392	4	US-09-073-009-138	Sequence 138, App
2	809	29.6	729	4	US-09-223-040-2	Sequence 2, Appli
3	737.5	26.9	585	4	US-09-620-412C-337	Sequence 337, App
4	737.5	26.9	585	4	US-09-598-419-337	Sequence 337, App
5	728	26.6	583	4	US-09-620-412C-353	Sequence 353, App
6	728	26.6	583	4	US-09-598-419-353	Sequence 353, App
7	722.5	26.4	518	4	US-09-620-412C-333	Sequence 333, App
8	722.5	26.4	518	4	US-09-598-419-333	Sequence 333, App
9	717.5	26.2	525	4	US-09-556-877-196	Sequence 196, App
10	717.5	26.2	525	4	US-09-620-412C-196	Sequence 196, App
11	717.5	26.2	525	4	US-09-598-419-196	Sequence 196, App
12	709.5	25.9	715	4	US-09-620-412C-321	Sequence 321, App
13	709.5	25.9	715	4	US-09-598-419-321	Sequence 321, App
14	708	25.9	715	4	US-09-620-412C-329	Sequence 329, App
15	708	25.9	715	4	US-09-598-419-329	Sequence 329, App
16	705	25.8	654	4	US-09-620-412C-341	Sequence 341, App
17	705	25.8	654	4	US-09-598-419-341	Sequence 341, App
18	704	25.7	723	4	US-09-736-457-1864	Sequence 1864, Ap
19	704	25.7	723	4	US-09-736-457-1864	Sequence 1864, Ap
20	704	25.7	691	4	US-09-620-412C-313	Sequence 313, App
21	704	25.7	691	4	US-09-598-419-313	Sequence 313, App
22	701.5	25.6	646	4	US-09-620-412C-317	Sequence 317, App
23	701.5	25.6	646	4	US-09-598-419-317	Sequence 317, App
24	699.5	25.6	700	4	US-09-620-412C-345	Sequence 345, App
25	699.5	25.6	700	4	US-09-598-419-345	Sequence 345, App
26	698	25.5	267	4	US-09-643-597-352	Sequence 352, App
27	698	25.5	267	4	US-09-606-421B-352	Sequence 352, App

ALIGNMENTS

RESULT 1

US-09-073-009-138
; Sequence 138, Application US/09073009

; Patent No. 6555653

; GENERAL INFORMATION:

; APPLICANT: Alderson, Mark

; APPLICANT: Dillon, Davin C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Campos-Neto, Antonio

; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF

; NUMBER OF SEQUENCES: 144

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY

; STREET: 6300 Columbia Center, 701 Fifth Ave.

; CITY: Seattle

; STATE: Washington

; COUNTRY: US

; ZIP: 98104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/073,009

; FILING DATE: 05-MAY-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.441C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-622-4900

; TELEFAX: 206-682-6031

; INFORMATION FOR SEQ ID NO: 138:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 392 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-073-009-138

Query Match 71.7%; Score 1963; DB 4; Length 392;

Best Local Similarity 99.7%; Pred. No. 4 5e-166;

Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 148 MSRAFIPTISAIDGLYDLIGIPNQGGILYSSLEYFEKALELEAAAPPGDWLGSA 207

Db 1 MSRAFIPTISAIDGLYDLIGIPNQGGILYSSLEYFEKALELEAAAPPGDWLGSA 60

TUBERCULOSIS AND MEI

Sequence 325, App
Sequence 325, App
Sequence 354, App
Sequence 354, App
Sequence 349, App
Sequence 349, App
Sequence 309, App
Sequence 309, App
Sequence 357, App
Sequence 357, App
Sequence 79, App
Sequence 80, App
Sequence 79, App
Sequence 80, App
Sequence 66, App
Sequence 67, App
Sequence 67, App

4 US-09-620-412C-325
4 US-09-598-419-325
4 US-09-643-597-354
4 US-09-606-421B-354
4 US-09-620-412C-349
4 US-09-598-419-349
4 US-09-620-412C-309
4 US-09-598-419-309
4 US-09-620-412C-357
4 US-09-598-419-357
3 US-08-818-112-79
3 US-08-818-111-80
3 US-09-056-556-79
4 US-09-072-596-80
3 US-08-818-112-66
3 US-08-818-111-67
4 US-09-056-556-66
4 US-09-072-596-67

28 697.5 25.5 631
29 697.5 25.5 631
30 696 25.4 299
31 696 25.4 299
32 696 25.4 487
33 696 25.4 487
34 696 25.4 619
35 696 25.4 619
36 696 25.4 683
37 696 25.4 683
38 671 24.5 355
39 671 24.5 355
40 671 24.5 355
41 671 24.5 355
42 670 24.5 132
43 670 24.5 132
44 670 24.5 132
45 670 24.5 132

```

QY 208 DKYAGKRNHNHVFQELADLDROLISLIHDQANAVOTTRDIIEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNHVFQELADLDROLISLIHDQANAVOTTRDIIEGAKKGLFVRPVAVDLT 120

QY 268 YIPVGHASAAQAFPCAGAMAVVGGALAYLVVKTLLINATOLLKLLAKLAEVAAAAD 327
Db 121 YIPVGHASAAQAFPCAGAMAVVGGALAYLVVKTLLINATOLLKLLAKLAEVAAAAD 180

QY 328 IISDVADIIGKILGEVWEFITTALNGLKELWDLKLTGWVTGLFSRGSNLSSEFFAGVPGLT 387
Db 181 IISDVADIIGKILGEVWEFITTALNGLKELWDLKLTGWVTGLFSRGSNLSSEFFAGVPGLT 240

QY 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASASLPALAGIGGSGFGGLPSLAQVHAA 300

QY 448 STRQALRPRADGPGVGAAGAEQVGSQSOLVSAQSGQMGPGVGMGHPSSGASKGTTTKY 507
Db 301 STRQALRPRADGPGVGAAGAEQVGSQSOLVSAQSGQMGPGVGMGHPSSGASKGTTTKY 360

QY 508 SEGAAAGTDAERAPVEADAGGQKYLVRNV 539
Db 361 SEGAAAGTDAERAPVEADAGGQKYLVRNV 392

RESULT 2
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Meto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ. ID NOS: 10
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion

US-09-223-040-2
Query Match 29.6%; Score 809; DB 4; Length 729;
Best Local Similarity 40.0%; Pred. No. 3.5e-63;
Matches 221; Conservative 69; Mismatches 170; Indels 92; Gaps 20;

QY 1 MHHHHHTAASDNFQISQSGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQISQSGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVVGSAFAASLGISTGDIVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 61 NNGNGARVQVVGSAFAASLGISTGDIVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGGTRTGNVTLAEGPPAEFLVPRGMSRAPIIDPTISALDGLYLLGIG-IP 173
Db 121 TKSGGTRTGNVTLAEGPPAEFLVPRGMSRAPIIDPTISALDGLYLLGIG-IP 173

QY 173 PNOGGILYSSLEFEKALEELAAAPGDCGLGSAADKYGKRNHNVFFQ-----ELAD 226
Db 175 -SVASDLFSAASAFQVVGWGLTVG-----SWIGSSAGLMVAASPIVAMSVTAGAELTA 229

QY 227 LDRQLISLIHDQANAVOTTRDIIEGAKKGLFVRPVAVDLT--TYIPVVG-----HA 275
Db 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILIATNLLQNTPAIAVNAEYGEWMA 287

```

```

QY 276 LSAAQFAFFCAGAMAVVGGALAYLVVKTLLINATOLLKLLAKLAEVAAAAD-IISDVAD 334
Db 288 QDAAMFGYAAATATATATLLPPEEAPEMTSAGGLEQAAAEEASPTAAANQLMNVVQ 347

QY 335 IIKOI-----LGEVWEFITTALNGLKELWDLKLTGWVTGLFSRGSNLSSEFFAG 382
Db 348 ALQOLAQTQGTTPSSKLGGLWKTVPSP-----ISNVVSMANNHMSMTNSG 396

QY 383 VPGLTGATSGLSQVTLFGAAG-----LSASSGLAHADSLASASLPALAG-----IGGG 432
Db 397 V-SWNTLS--SMKGFAPAARQAVQTAQNGVRAMSSLGSSLGSSGLGGVAAANLGRA 453

QY 433 SGFGGLPSLAQVHAASTRQALRPRADG-----PVGAAAEQVGSQSOLVSAQSGQMG- 487
Db 454 ASVGSGL-SVPOAMAAA-NQAVTPAARALPLTSLTSAERPGQW-----LGLLPV 501

QY 488 GMGGMHPSSGAS 499
Db 502 QMGARAGGGLS 513

RESULT 3
US-09-620-412C-337
; Sequence 337, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-337
Query Match 26.9%; Score 737.5; DB 4; Length 585;
Best Local Similarity 38.8%; Pred. No. 5.5e-57;
Matches 231; Conservative 40; Mismatches 174; Indels 151; Gaps 22;

QY 1 MHHHHHTAASDNFQISQSGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQISQSGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 56

QY 61 NNGNGARVQVVGSAFAASLGISTGDIVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 57 NNGNGARVQVVGSAFAASLGISTGDIVTAVDGPINSATAMADALNGHHPGDVSVTWQ 116

QY 121 TKSGGTRTGNVTLAEGPPAEFLVPRGMSRAPIIDPTISALDGLYLLGIG-IP 173
Db 117 TKSGGTRTGNVTLAEGPPAEFLVPRGMSRAPIIDPTISALDGLYLLGIG-IP 173

QY 174 NOGGILYSSLE-----YFEKA-----LEELAAAPGDCGLGSAADKYGKRNH 217
Db 174 VEGGINFQDLEETRIKYNKAGTETKKTILPSLKAQASAGNADAWASSPQSGS----- 228

QY 218 VNFQELADLDROLISLIHDQANAVOTTRDIIEGAKKGLFVRPVAV-----DLTY 268
Db 229 -----ATTVSDGSDSSGSDS*SETVPVTAAGGGLYTDKNLSI 267

QY 269 IPVVGHALSAFQAFPCAGAMAVVGGALAYLVVKTLLINATOLLKLLAKLAEVAAAAD 328
Db 269 TNITG-IIEIANNKATDVGGGAVVKGT-----TCENSHR-LQFLKNSDKQGG----- 314

QY 329 ISDVADIIGKILGEVWEFITTALNGLKELWDLKLTGWVT--GLFSRG-----WSNLEST- 379
Db 315 -----GIYGEDNITLSN-LTG-KTLFQENTAKEBGGGLFKGTDXALTWTGLDLSFC 363

```

```
Qy 380 -----FAGVPLGTGATSGLSQVGTG---LFGAAGLSASSG 410
Db 364 LINNTSEKHGGGAFVTKETISQYTSVDETIPGIT-PVHGETVITGNKSTGGNGGVCYTKR 422
Qy 411 LAHADSLASSALPALAGTGGGSGGGLPSLAQVHAASRQALRPADCPVCAAAEQVGG 470
Db 423 LALSNIQSISISGNSAAENGSG-----AHTCPDSFFTAD-----TAEQPA 463
Qy 471 QSQVLSAQSGGMPGVGMGHPSSGASKGTTTKYSEGAAAGTDAERAPVEAD 526
Db 464 ASAATSTPKS-----APVSTALSTPSSSTVSSLLTAAASSQASPATSKETQDPNAD 515

RESULT 4
US-09-598-419-337
; Sequence 337, Application US/09598419
; Patent No. 6565856
; GENERAL INFORMATION:
; APPLICANT: Scheik, Yasir A.W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C6
; CURRENT APPLICATION NUMBER: US/09/598,419
; CURRENT FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 357
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-598-419-337

Query Match 26.9%; Score 737.5; DB 4; Length 585;
Best Local Similarity 38.8%; Pred. No. 5.5e-57;
Matches 231; Conservative 40; Mismatches 174; Indels 151; Gaps 22;

Qy 1 MHHHHHTAASDNFQSQGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVD 56

61 NNGNGARVQVVGSAAPASLGISTGVTAVDGPINSATAMADALNGHHPGDVISTWQ 120
57 NNGNGARVQVVGSAAPASLGISTGVTAVDGPINSATAMADALNGHHPGDVISTWQ 116

121 TKSQGTTRGNVTLAEGPPAEFL-----VPRGMSRAFIIDPTISADGLYDLIG-IP 173
117 TKSQGTTRGNVTLAEGPPAEFCRYPSHWRPLDGLS---IQSKQSLFNSYKQGGALY 173

174 NOGGILYSSLE-----YFEKA-----LEELAAFPDGLWLSAADKYAGKRNH 217
174 VEGGINFQDLERIRIKYNKAGTFETKITLPSKAQAGNADAWASSPQSGG----- 228

218 VNFQBELADLRQLISLIHQANAVQTTDILEGAKKLEFVRPVAV-----DLTY 268
229 -----ATTVSDGSSGSDSTSETVPTAKGGLYTDKNLSI 267

269 IPVGHALSAAPQAPFCAGAMAVGGALAYLVKTLINATQLKLAELVAAAIADI 328
268 TMITG-TIIBIANNKATDVGGGAYVKGTL-----TCENSHR-LQFLKNSDKQGG----- 314

329 ISDVADIKGLGEVWEFTNGLKELWDKLTGWT--GLFSRG-----WSNLESP- 379
315 -----GIYGEDNITLSN-LTG-KTLFQENTAKEGGGLFIKGTDKALTMGLDSCF 363

380 -----FAGVPLGTGATSGLSQVGTG---LFGAAGLSASSG 410
Db 364 LINNTSEKHGGGAFVTKETISQYTSVDETIPGIT-PVHGETVITGNKSTGGNGGVCYTKR 422
Qy 411 LAHADSLASSALPALAGTGGGSGGGLPSLAQVHAASRQALRPADCPVCAAAEQVGG 470
Db 423 LALSNIQSISISGNSAAENGSG-----AHTCPDSFFTAD-----TAEQPA 463
```

```
Qy 471 QSQVLSAQSGGMPGVGMGHPSSGASKGTTTKYSEGAAAGTDAERAPVEAD 526
Db 464 ASAATSTPKS-----APVSTALSTPSSSTVSSLLTAAASSQASPATSKETQDPNAD 515

RESULT 5
US-09-620-412C-353
; Sequence 353, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 353
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-353

Query Match 26.6%; Score 728; DB 4; Length 583;
Best Local Similarity 37.8%; Pred. No. 3.8e-56;
Matches 228; Conservative 49; Mismatches 186; Indels 140; Gaps 21;

Qy 1 MHHHHHTAASDNFQSQGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVD 56

61 NNGNGARVQVVGSAAPASLGISTGVTAVDGPINSATAMADALNGHHPGDVISTWQ 120
57 NNGNGARVQVVGSAAPASLGISTGVTAVDGPINSATAMADALNGHHPGDVISTWQ 116

121 TKSQGTTRGNVTLAEGPPAEF-----LVPRGMSRAFIIDPTISADGLYDLIG 171
117 TKSQGTTRGNVTLAEGPPAEFCRYPSHWRPLDQVSESPSTPSD-----DVLGKG 168

172 IPNCGGILYSSLEFYFEKALEELAAFPDGLWLSAADKYAG-----KRNHNVNFFQ 222
169 ----GGI-----YTEKSLTITGITITDFVSNIAIDSGAGVFTKENLCTNTNSLQFLK 218

223 ELADLD-----RQLISLIHQANAVQTTDIL--EGAKKG-----LEFVR 260
219 NSAGQHGGGAYVTQTHSVTNTTSETITPPLVGEVIFSENTAKHGCGGICTNKLISLNK 278

261 PVAV-----DLTYIPVVGHALSAAPQAPFCAGAMAVGGALAYLVKTLIN 306
279 TVTLTKNSAKESGGAIFTDLASITPTDTPESSTPSSSPASTPEVVASA----- 327

307 ATOLLKILAKLAELVAAAIADIISDVAD-----IIKGLGEVWEFTNGLNG 353
328 --KINRFFASTAPAPSLTEAESDQTDQTDTSNDSIDVSIENILNVAINQNTSAKKG 395

354 LKELWDKLTGWTGLFSR--GWSNLESPFAGVPLGTGATS-----GLSQVTGL-FGAAGL 405
386 -----CAIYKXKAKLSRINNLE-----LSGNSQDVGGGLCLTESVEFDAIG- 427

406 SASGLAHADSLASSASL-----PALAGICGGSGGGLPSLAQVHAASRQALRPADGP 460
428 ---SLUSHYNSAAKEGQVHKSQVTLNLSKSTTFADNTVKAIVESTPEAPEEIPVBE 484
461 VAAAAE---QVGGQSOLVSAQSGQMGQPVGMGHPSSGASKGTTTKYSEGAAAGTE 516
485 ESTATENPNNTGSSANTNLEGSQDGTADTGTGVVNNESQDTSDTGNAESGQLQDSIQ 544

517 DAE 519
545 SNE 547
```

RESULT 6

US-09-598-419-353

; Sequence 353, Application US/09598419

; Patent No. 6565856

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Scholler, John

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.469C6

; CURRENT APPLICATION NUMBER: US/09/598,419

; CURRENT FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 357

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 353

; LENGTH: 583

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-09-598-419-353

Query Match 26.6%; Score 728; DB 4; Length 583;

Best Local Similarity 37.8%; Pred. No. 3.8e-56;

Matches 228; Conservative 49; Mismatches 186; Indels 140; Gaps 21;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMATAGQIRSGGSPVTHIGTAFGLGVVD 60

Db 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMATAGQIK-----LPTVHIGTAFGLGVVD 56

QY 61 NNGNGARVQVVGSAAPASLGISTGDVITAVDGAPINSATAMADALNGHHPGDIVSVTWQ 120

Db 57 NNGNGARVQVVGSAAPASLGISTGDVITAVDGAPINSATAMADALNGHHPGDIVSVTWQ 116

QY 121 TKSGETRTGNVTLAEGPPAPF-----LVPRGSMRAFIIDPTISAIIDGLYDLGI 171

Db 117 TKSGETRTGNVTLAEGPPAPFCRYPHWRPLMKWLSATAVFAAIVLPSV-----GF 167

QY 172 IPNQGGILYSLEFYEFKALEELAAFPDGLWLSAADKVAG-----KRNHNVNFFQ 222

Db 169 -----GGI-----YTKSLIITIGTIDFVSNIAIDSGAVFTKENLCTWNSIQFLK 218

QY 223 ELADL-----ROLISLHDAQNAVQ-----TRDIL--EGAKG-----LEFVR 260

Db 219 NSAGQGGGAYVTQWTSVNTTSESITPLVGEVIFSENTAKHGCGGICTNKLSSLNLK 278

QY 261 PVAV-----DUTYIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLLN 306

Db 279 TVLTKNKAKESGGAFTDLASITPTDTPESSTPSSSPASTPEVVASA-----327

QY 307 ATOLLKLAKLAELVAAAADIISDVAD-----IIGKILGEVWEFITTALNG 353

Db 328 --KINRFFASTAPAPASLSTEAESDQTDQTDTSNDSIDVSIENILVAINONTSAKKG 385

QY 354 LKELWDLKLTGVTGLFSR--GWSNLSFFAGVPGLTGATS-----GLSQVTGL--FGAAGL 405

Db 386 -----GAIYKXKAKUGRINLLE-----LSGNSQDVGGGLCTESVEFPAIG- 427

QY 406 SASGLAHADSLASSASL-----PALAGIGGGGFGGLPSLAQVHAASRQALRPADGP 460

Db 428 ---SLLSHYNSAAKEGGVIHSKVTLSNLKSTFTFADNTVKAIVESTPEAPEIIPVEGE 484

QY 461 VGRAAE-----QVGGQSLVSAQSQGGGPGVGMGHPSPSGASKGTTTKYSEGAAGTE 516

Db 485 ESTATENPNSTEGSSANTNLEGSQDGTADTGTGVVNNESQDTSDTGNAESGEQLQDSTQ 544

QY 517 DAE 519

Db 545 SNE 547

RESULT 7

US-09-620-412C-333

; Sequence 333, Application US/09620412C

; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Flind

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 333

; LENGTH: 518

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-09-620-412C-333

Query Match 26.4%; Score 722.5; DB 4; Length 518;

Best Local Similarity 37.3%; Pred. No. 9.9e-56;

Matches 209; Conservative 44; Mismatches 132; Indels 175; Gaps 21;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMATAGQIRSGGSPVTHIGTAFGLGVVD 60

Db 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMATAGQIK-----LPTVHIGTAFGLGVVD 56

QY 61 NNGNGARVQVVGSAAPASLGISTGDVITAVDGAPINSATAMADALNGHHPGDIVSVTWQ 120

Db 57 NNGNGARVQVVGSAAPASLGISTGDVITAVDGAPINSATAMADALNGHHPGDIVSVTWQ 116

QY 121 TKSGETRTGNVTLAEGPPAPF-----LVPRGSMRAFIIDPTISAIIDGLYDLGI 170

Db 117 TKSGETRTGNVTLAEGPPAPFCRYPHWRPLMKWLSATAVFAAIVLPSV-----GF 167

QY 171 GIPNQGGILYSLEFYEFKALEELAAFPDGLWLG-----SAAKYAGKNHNV 218

Db 168 CFPPEKELNFRVETSSSTFTTETIGAGAGYIVSGNASFTKTNITPTTTTPTNSNS 227

QY 219 NFOELADLDRQLISLHDAQNAVQTTDILEGAKGLEFVRPVAVDLTIPVVGHALSA 278

Db 228 SSSGETAS-----VSESDSDSTTTTPD-----PKG-----251

QY 279 AFQAPFCAGAMAVVGGALAYLVVKTLLNATQLLKLAKLAELVAAAADIISDVADIIG 338

Db 252 -----GGA-----FYNH-----SG 261

QY 339 ILGEVWEFITTALNGKELMDKLTGVTGLFSPRGSWNLSFFAGVPGLTGATSLQVTG 398

Db 262 VLSFWTRSGTSGSLTSEI--KMTGEGGAIFSG-----ELLFTDLTSLT-IQNNLSQLSG 314

QY 399 --LFGAAGLSASSGLAHADSLASSASL--PALAGIGGGGFGGLPSLAQVHAASRQALRP 456

Db 315 GAIFGGSTISL-SGITKATFSCNSAEP-----APVKKPTPEK 351

QY 457 ADGPVGAAGVGGQSLVSAQSQGGGPGVGMGHPSPSGASKGTTTKYSE-----G 510

Db 352 AQ-----TASTSSSS-----SSGNSDVS-----PSSRAEPAAANLQSHFICATAT 395

QY 511 AAAGTEDAERAPVADAGG 530

Db 396 PAAQTDSTSTSPSHKPGSGG 415

RESULT 8

US-09-598-419-333

; Sequence 333, Application US/09598419

; Patent No. 6565856

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Scholler, John

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.469C6

; CURRENT APPLICATION NUMBER: US/09/598,419

; CURRENT FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 357

```

; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 333
; LENGTH: 518
; TYPE: PRG
; ORGANISM: Chlamydia trachomatis
US-09-598-419-333

Query Match      26.4%; Score 722.5; DB 4; Length 518;
Best Local Similarity 37.3%; Pred. No. 9.9e-56;
Matches 209; Conservative 44; Mismatches 132; Indels 175; Gaps 21;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAAPASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 116
Qy 121 TKSGETRTGNVTLAEGPPAEF-LVPRGSMRAFIDPTISADGLYDLGLGIPNQGGL 179
Db 117 TKSGETRTGNVTLAEGPPAEF-LVPRGSMRAFIDPTISADGLYDLGLGIPNQGGL 157
Qy 180 YSLEYFEKALEEELAAAFPGDGLGSAAD 208
Db 158 -----IDGTWEGASGD 169

RESULT 10
US-09-620-412C-196
; Sequence 196, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 196
; LENGTH: 525
; TYPE: PRG
; ORGANISM: Chlamydia
US-09-620-412C-196

Query Match      26.2%; Score 717.5; DB 4; Length 525;
Best Local Similarity 71.3%; Pred. No. 2.8e-55;
Matches 149; Conservative 6; Mismatches 13; Indels 41; Gaps 4;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAAPASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 116
Qy 121 TKSGETRTGNVTLAEGPPAEF-LVPRGSMRAFIDPTISADGLYDLGLGIPNQGGL 179
Db 117 TKSGETRTGNVTLAEGPPAEF-LVPRGSMRAFIDPTISADGLYDLGLGIPNQGGL 157
Qy 180 YSLEYFEKALEEELAAAFPGDGLGSAAD 208
Db 158 -----IDGTWEGASGD 169

RESULT 11
US-09-598-419-196
; Sequence 196, Application US/09598419
; Patent No. 6565856
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Scholler, John
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C6
; CURRENT FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 196
; LENGTH: 525
; TYPE: PRG
; ORGANISM: Chlamydia
```

```
; CURRENT APPLICATION NUMBER: US/09/598,419
; CURRENT FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 357
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 196
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-598-419-196

Query Match      26.2%; Score 717.5; DB 4; Length 525;
Best Local Similarity 71.3%; Pred. No. 2.8e-55;
Matches 149; Conservative 6; Mismatches 13; Indels 41; Gaps 4;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHHPGDVISVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHHPGDVISVTWQ 116
Qy 121 TKSGETRTGNVTLAEGPPAEF-LVPRGMSRAFIIDPTISAIDGLYDLLGIGIPNQGIL 179
Db 117 TKSGETRTGNVTLAEGPPAEF-LVPRGMSRAFIIDPTISAIDGLYDLLGIGIPNQGIL 157
Qy 180 YSSLEVEFEKALEBAAAFDGDHGLSAAD 208
Db 158 -----IDGTMWEGASGD 169

RESULT 12
US-09-620-412C-321
; Sequence 321, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210421.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 321
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-321

Query Match      25.9%; Score 709.5; DB 4; Length 715;
Best Local Similarity 37.5%; Pred. No. 2.3e-54;
Matches 214; Conservative 53; Mismatches 187; Indels 117; Gaps 21;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHHPGDVISVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHHPGDVISVTWQ 116
Qy 121 TKSGETRTGNVTLAEGPPAEF-----LVPRGMSRA--FIIDPTISAIDGLYDLLG 169
Db 117 TKSGETRTGNVTLAEGPPAEFCRYPHWRPLDIRTLNGKEHNYIKEAPTTLKFGT----- 171
Qy 170 IGIPNQGILYSSLEVEFEKALEBAAAFDGDGMLGAADKYAGKRNHNVNFQELADLDR 229
Db 172 LAIEDDAEIEFNIPFQNTPTSLA-----LGSATLTGK----- 207
Qy 230 QLISLIHQANAVQ--TTDRILEGAKKGLFVRPVAVDLTYIPVVGHALS--AAFOAPFC 285
Db 208 -----HGKLNITNLGVILPIILKEGKSPPCIRVNPQDMTQNTGTGTPSSSTISSTP-- 259
Qy 286 AGAMAVVGGALAYLVKTLINATQLLAKLAEVAAAIADIISDVADIKILGEVWE 345
Db 260 ---MIENGRLS-IVDENYESVDSMDLSRGKAE-----QLILSIETTNDQOLDSNWQ 308
Qy 346 FITNALN-----GLKELWDKLTGWVTGLPSRGWSNLESFFAGVPGITGATSGLSQV 396
Db 309 ---SSLNTLSLSPPHYGYQGLW--TPNWITTTITILNNSS-----APTSATSAEQ-- 356

Query Match      25.9%; Score 709.5; DB 4; Length 715;
Best Local Similarity 37.5%; Pred. No. 2.3e-54;
Matches 214; Conservative 53; Mismatches 187; Indels 117; Gaps 21;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHHPGDVISVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHHPGDVISVTWQ 116
Qy 121 TKSGETRTGNVTLAEGPPAEF-----LVPRGMSRA--FIIDPTISAIDGLYDLLG 169
Db 117 TKSGETRTGNVTLAEGPPAEFCRYPHWRPLDIRTLNGKEHNYIKEAPTTLKFGT----- 171
Qy 170 IGIPNQGILYSSLEVEFEKALEBAAAFDGDGMLGAADKYAGKRNHNVNFQELADLDR 229
Db 172 LAIEDDAEIEFNIPFQNTPTSLA-----LGSATLTGK----- 207
Qy 230 QLISLIHQANAVQ--TTDRILEGAKKGLFVRPVAVDLTYIPVVGHALS--AAFOAPFC 285
Db 208 -----HGKLNITNLGVILPIILKEGKSPPCIRVNPQDMTQNTGTGTPSSSTISSTP-- 259
Qy 286 AGAMAVVGGALAYLVKTLINATQLLAKLAEVAAAIADIISDVADIKILGEVWE 345
Db 260 ---MIENGRLS-IVDENYESVDSMDLSRGKAE-----QLILSIETTNDQOLDSNWQ 308
Qy 346 FITNALN-----GLKELWDKLTGWVTGLPSRGWSNLESFFAGVPGITGATSGLSQV 396
Db 309 ---SSLNTLSLSPPHYGYQGLW--TPNWITTTITILNNSS-----APTSATSAEQ-- 356
```


QY 397 TGLFGAAGLSASGLAHADSLASSALP---ALAGIGGGGFGGLPSLAQVHAASTROAL 453
Db 357 -----KKTSETFTPSNTTASIPNIKASAGSGSAGNSGEVITTKHTLVNWA- 405
QY 454 RPRADGPVGAAGAEQVGGQQLYSAGSQSGGPGVGMGHPSS-----GASKGTTTKY 507
Db 406 -----PVGIVYDPIRRGLDIANSVLHSGRNMTWGLRSLPDSNFWALQGAATTLFTKQ 459
QY 508 SEGAAAGTDEAERA-PVEADAGG--GQKVLV 535
Db 460 KRLSVHGYSSAGKGYTVSSQASGAHGKFL 490

RESULT 14

US-09-620-412C-329

; Sequence 329, Application US/09620412C

; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Fling

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; SOFTWARE: FastSEQ for Windows Version 3.0/4.0

; SEQ ID NO 329

; LENGTH: 715

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-09-620-412C-329

Query Match 25.9%; Score 708; DB 4; Length 715;
Best Local Similarity 70.7%; Pred. No. 3.1e-54;
Matches 152; Conservative 13; Mismatches 26; Indels 24; Gaps 6;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGLVVD 60
Db 1 MHHHHHTAASDNFOLSGGGGFAIPGQAMAIAGQIK-----LPTVHIGPTAFGLGLVVD 56
QY 61 NNGNGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVSVTWQ 116
QY 121 TKSGETRTGNVTLAEGPPAEF-----LVPRGMSRAFIIDPTISAIDGLYDLL 168
Db 117 TKSGETRTGNVTLAEGPPAEFCRYPHWRPLDPVVQNSAAGASTPSSSSMPG---AV 173
QY 169 GIGIPNOGGILYS--SLEYFEKALEEL--AAAPFG 199
Db 174 TINQSGNSVIFTAESLTPSEK-LQVLNSTSNFPG 207

RESULT 15

US-09-598-419-329

; Sequence 329, Application US/09598419

; Patent No. 6565856

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Scholler, John

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C6

; CURRENT APPLICATION NUMBER: US/09/598,419

; CURRENT FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 357

; SOFTWARE: FastSEQ for Windows Version 3.0/4.0

; SEQ ID NO 329

; LENGTH: 715

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-09-598-419-329

Query Match 25.9%; Score 708; DB 4; Length 715;
Best Local Similarity 70.7%; Pred. No. 3.1e-54;
Matches 152; Conservative 13; Mismatches 26; Indels 24; Gaps 6;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGLVVD 60
Db 1 MHHHHHTAASDNFOLSGGGGFAIPGQAMAIAGQIK-----LPTVHIGPTAFGLGLVVD 56
QY 61 NNGNGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVSVTWQ 116
QY 121 TKSGETRTGNVTLAEGPPAEF-----LVPRGMSRAFIIDPTISAIDGLYDLL 168
Db 117 TKSGETRTGNVTLAEGPPAEFCRYPHWRPLDPVVQNSAAGASTPSSSSMPG---AV 173
QY 169 GIGIPNOGGILYS--SLEYFEKALEEL--AAAPFG 199
Db 174 TINQSGNSVIFTAESLTPSEK-LQVLNSTSNFPG 207

Search completed: November 21, 2003, 16:11:44

Job time : 12.3667 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 21.0222 Seconds

(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-64

Perfect score: 2737

Sequence: 1 MHHHHHTAASDNFQLSQGG.....RAPVEADAGGQKVLVRNV 539

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09D_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1963	71.7	392	9	US-09-073-009-138
2	1963	71.7	392	9	US-09-793-306-138
3	1963	71.7	392	12	US-10-098-732A-29
4	1963	71.7	392	16	US-10-080-170-610
5	1263	45.1	394	16	US-10-080-170-53
6	809	29.6	729	9	US-09-287-849-2
7	809	29.6	729	12	US-10-359-460-2
8	809	29.6	729	12	US-10-098-732A-16
9	809	29.6	729	12	US-10-098-732A-18
10	809	29.6	930	12	US-10-098-732A-65
11	739	27.0	231	9	US-09-287-849-28
12	739	27.0	231	12	US-10-359-460-28
13	737.5	26.9	585	9	US-09-841-132-337
14	728	26.6	583	9	US-09-841-132-353
15	722.5	26.4	518	9	US-09-841-132-333

16	718.5	26.3	529	9	US-09-810-936-324
17	718.5	26.3	529	10	US-09-924-400-324
18	718.5	26.3	529	15	US-10-213-679-324
19	717.5	26.2	525	9	US-09-841-132-196
20	709.5	25.9	715	9	US-09-841-132-321
21	708	25.9	715	9	US-09-841-132-329
22	706.5	25.8	384	10	US-09-902-941-1876
23	706.5	25.8	384	10	US-09-849-626-1876
24	706.5	25.8	384	12	US-10-113-872-1876
25	706.5	25.8	384	15	US-10-017-754-1876
26	705.5	25.8	344	9	US-09-922-217-1085
27	705.5	25.8	344	10	US-09-833-263-1085
28	705.5	25.8	344	14	US-10-025-380-1085
29	705	25.8	654	9	US-09-841-132-341
30	704	25.7	273	10	US-09-736-457-1864
31	704	25.7	273	10	US-09-902-941-1864
32	704	25.7	273	10	US-09-849-626-1864
33	704	25.7	273	12	US-10-113-872-1864
34	704	25.7	273	15	US-10-017-754-1864
35	704	25.7	314	10	US-09-736-457-1863
36	704	25.7	314	10	US-09-902-941-1863
37	704	25.7	314	10	US-09-849-626-1863
38	704	25.7	314	12	US-10-113-872-1863
39	704	25.7	314	15	US-10-017-754-1863
40	704	25.7	400	9	US-09-759-143-852
41	704	25.7	400	9	US-09-780-669-852
42	704	25.7	400	9	US-09-822-827-852
43	704	25.7	400	10	US-09-895-793-852
44	704	25.7	400	10	US-09-895-814-852
45	704	25.7	400	12	US-10-144-678A-852

ALIGNMENTS

RESULT 1

US-09-073-009-138
; Sequence 138, Application US/09073009
; Patent No. US20010012888A1
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid

TUBERCULOSIS AND ME

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-073-306-138

Query Match
Best Local Similarity 99.7%; Score 1963; DB 9; Length 392;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 148 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAFPDGLGSA 207
Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAFPDGLGSA 60

Qy 208 DKYAGKRNHNVPFQELADLDROLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 267
Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAFPDGLGSA 60

Qy 208 DKYAGKRNHNVPFQELADLDROLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNVPFQELADLDROLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 120

Qy 268 YIPVVGHALSAAPQAPFCAGAMAVGGALAYLVVKTLLINATQLLLAKLAELVAAAIAD 327
Db 1 YIPVVGHALSAAPQAPFCAGAMAVGGALAYLVVKTLLINATQLLLAKLAELVAAAIAD 180

Qy 328 IISDVADIIGKILGEVWEFITNALNGELKELWDKLTGWVTGLFSRGSNSLESFFAGVPGLT 387
Db 181 IISDVADIIGKILGEVWEFITNALNGELKELWDKLTGWVTGLFSRGSNSLESFFAGVPGLT 240

Qy 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300

Qy 448 STQALRPADGPGVAAAQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKKY 507
Db 301 STQALRPADGPGVAAAQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKKY 360

Qy 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 2
US-09-793-306-138
; Sequence 138, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Ovendale, Pamela
; APPLICANT: Jen, Shyan
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 138
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: hTcc#1
US-09-793-306-138

Query Match
Best Local Similarity 99.7%; Score 1963; DB 9; Length 392;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 148 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAFPDGLGSA 207
Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAFPDGLGSA 60

Qy 208 DKYAGKRNHNVPFQELADLDROLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNVPFQELADLDROLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 120

Qy 268 YIPVVGHALSAAPQAPFCAGAMAVGGALAYLVVKTLLINATQLLLAKLAELVAAAIAD 327
Db 121 YIPVVGHALSAAPQAPFCAGAMAVGGALAYLVVKTLLINATQLLLAKLAELVAAAIAD 180

Qy 328 IISDVADIIGKILGEVWEFITNALNGELKELWDKLTGWVTGLFSRGSNSLESFFAGVPGLT 387
Db 181 IISDVADIIGKILGEVWEFITNALNGELKELWDKLTGWVTGLFSRGSNSLESFFAGVPGLT 240

Qy 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300

Qy 448 STQALRPADGPGVAAAQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKKY 507
Db 301 STQALRPADGPGVAAAQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKKY 360

Qy 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 3
US-10-098-732A-29
; Sequence 29, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB40 (HTCC#1)
US-10-098-732A-29

Query Match
Best Local Similarity 99.7%; Score 1963; DB 12; Length 392;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 148 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAFPDGLGSA 207
Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAFPDGLGSA 60

Qy 208 DKYAGKRNHNVPFQELADLDROLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNVPFQELADLDROLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 120

Qy 268 YIPVVGHALSAAPQAPFCAGAMAVGGALAYLVVKTLLINATQLLLAKLAELVAAAIAD 327
Db 121 YIPVVGHALSAAPQAPFCAGAMAVGGALAYLVVKTLLINATQLLLAKLAELVAAAIAD 180

Qy 328 IISDVADIIGKILGEVWEFITNALNGELKELWDKLTGWVTGLFSRGSNSLESFFAGVPGLT 387
Db 181 IISDVADIIGKILGEVWEFITNALNGELKELWDKLTGWVTGLFSRGSNSLESFFAGVPGLT 240
```

QY 388 GATGSLSVTGLFAGAGLSASGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
 |||||
 Db 241 GATGSLSVTGLFAGAGLSASGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 |||||
 QY 448 STQALRRADPGVGAARAEQVGGQSOLVSAOGSCMGPGVGMGHPSSGASKGTTTKY 507
 |||||
 Db 301 STQALRRADPGVGAARAEQVGGQSOLVSAOGSCMGPGVGMGHPSSGASKGTTTKY 360
 |||||
 QY 508 SEGAAAGTEDAERAPVEADAGGQKVLVRNVV 539
 |||||
 Db 361 SEGAAAGTEDAERAPVEADAGGQKVLVRNVV 392
 |||||

RESULT 4

US-10-080-170-610
 ; Sequence 610, Application US/10080170
 ; Publication No. US20030129601A1

GENERAL INFORMATION:
 ; APPLICANT: COLE, S.T.
 ; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
 ; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
 ; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
 ; FILE REFERENCE: 03495.0218
 ; CURRENT APPLICATION NUMBER: US/10/080,170
 ; PRIOR FILING DATE: 2002-06-10
 ; PRIOR FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 652
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 610
 ; LENGTH: 392
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-10-080-170-610

Query Match 71.7%; Score 1963; DB 16; Length 392;
 Best Local Similarity 99.7%; Pred. No. 4.9e-145;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPGDMGWSAA 207
 |||||
 Db 1 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPGDMGWSAA 60
 |||||
 QY 208 DKYAGKRNHNHVFQELADLRQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 267
 |||||
 Db 61 DKYAGKRNHNHVFQELADLRQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 120
 |||||
 QY 268 YIPVVGHALSAAPFQACAGAMAVGGALAYLVVKTLLINATQLKLAKLAELVAAAIAD 327
 |||||
 Db 121 YIPVVGHALSAAPFQACAGAMAVGGALAYLVVKTLLINATQLKLAKLAELVAAAIAD 180
 |||||
 QY 328 IISDVADIKIGLGEVWEFFITNALNGKELDKLTGWVTGLFSRGNLSFFAGVPGLT 387
 |||||
 Db 181 IISDVADIKIGLGEVWEFFITNALNGKELDKLTGWVTGLFSRGNLSFFAGVPGLT 240
 |||||
 QY 388 GATGSLSVTGLFAGAGLSASGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
 |||||
 Db 241 GATGSLSVTGLFAGAGLSASGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 |||||
 QY 448 STQALRRADPGVGAARAEQVGGQSOLVSAOGSCMGPGVGMGHPSSGASKGTTTKY 507
 |||||
 Db 301 STQALRRADPGVGAARAEQVGGQSOLVSAOGSCMGPGVGMGHPSSGASKGTTTKY 360
 |||||

RESULT 5

US-10-080-170-53
 ; Sequence 53, Application US/10080170
 ; Publication No. US20030129601A1

GENERAL INFORMATION:

; APPLICANT: COLE, S.T.
 ; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
 ; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
 ; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
 ; FILE REFERENCE: 03495.0218
 ; CURRENT APPLICATION NUMBER: US/10/080,170
 ; CURRENT FILING DATE: 2002-06-10
 ; PRIOR APPLICATION NUMBER: 60/270,123
 ; PRIOR FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 652
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 53
 ; LENGTH: 394
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium leprae
 US-10-080-170-53

Query Match 46.1%; Score 1263; DB 16; Length 394;
 Best Local Similarity 62.9%; Pred. No. 2.1e-90;
 Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;

QY 148 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPGDMGWSAA 207
 |||||
 Db 1 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPGDMGWSAA 60
 |||||
 QY 208 DKYAGKRNHNHVFQELADLRQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 267
 |||||
 Db 61 DKYAGKRNHNHVFQELADLRQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 120
 |||||
 QY 268 YIPVVGHALSAAPFQACAGAMAVGGALAYLVVKTLLINATQLKLAKLAELVAAAIAD 327
 |||||
 Db 121 YIPVVGHALSAAPFQACAGAMAVGGALAYLVVKTLLINATQLKLAKLAELVAAAIAD 180
 |||||
 QY 328 IISDVADIKIGLGEVWEFFITNALNGKELDKLTGWVTGLFSRGNLSFFAGVPGLT 387
 |||||
 Db 181 IISDVADIKIGLGEVWEFFITNALNGKELDKLTGWVTGLFSRGNLSFFAGVPGLT 240
 |||||
 QY 388 GATGSLSVTGLFAGAGLSASGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
 |||||
 Db 241 GATGSLSVTGLFAGAGLSASGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 |||||
 QY 448 STQALRRADPGVGAARAEQVGGQSOLVSAOGSCMGPGVGMGHPSSGASK--GTTTK 505
 |||||
 Db 301 STQALRRADPGVGAARAEQVGGQSOLVSAOGSCMGPGVGMGHPSSGASK--GTTTK 360
 |||||
 QY 506 KYSEGAAGTEDAERAPVEADAGGQKVLVRNVV 539
 |||||
 Db 361 KYSEGAAGTEDAERAPVEADAGGQKVLVRNVV 394
 |||||

RESULT 6

US-09-287-849-2
 ; Sequence 2, Application US/09287849
 ; Patent No. US20020003459A1

GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; TITLE OF INVENTION: and Their Uses
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/09/287,849
 ; CURRENT FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match 29.6%; Score 809; DB 9; Length 729;
Best Local Similarity 40.0%; Pred. No. 1.4e-54;
Matches 221; Conservative 69; Mismatches 170; Indels 92; Gaps 20;

QY 1 MHHHHHTAASDNFQSQGGGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAFASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 120
Db 61 NNGNGARVQVVGSAFASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 120
QY 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLIGIGI 172
Db 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLIGIGI 172
QY 173 PNQGGILYSSLEFEXKALEELAAAFPCDGLGSAADKYAGKNRNHVNFFQ-----ELAD 226
Db 173 PNQGGILYSSLEFEXKALEELAAAFPCDGLGSAADKYAGKNRNHVNFFQ-----ELAD 226
QY 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSSAGLMVAASPYVAMSVTAGQALTA 229
Db 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSSAGLMVAASPYVAMSVTAGQALTA 229
QY 227 LDRQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
Db 227 LDRQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
QY 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILITATLLGQNTPAIAVNEAEYGEWMA 287
Db 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILITATLLGQNTPAIAVNEAEYGEWMA 287
QY 276 LSAAFOAPFCAGAVAVVGGALAYLVVKTLINATQLLKLAKLAELVAAAIAD--IISDVAD 334
Db 276 LSAAFOAPFCAGAVAVVGGALAYLVVKTLINATQLLKLAKLAELVAAAIAD--IISDVAD 334
QY 433 SGFGGLPSLAQVHAASRQALRPRADG-----PVGAAAEQVGGQSGLVSAQSGQMGG-PV 487
Db 433 SGFGGLPSLAQVHAASRQALRPRADG-----PVGAAAEQVGGQSGLVSAQSGQMGG-PV 487
QY 488 GMGCMHPSSGAS 499
Db 502 GQMGARAGGGLS 513

RESULT 7

US-10-359-460-2
; Sequence 2, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and their Uses
; FILE REFERENCE: 014058-009020US

CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match 29.6%; Score 809; DB 12; Length 729;
Best Local Similarity 40.0%; Pred. No. 1.4e-54;
Matches 221; Conservative 69; Mismatches 170; Indels 92; Gaps 20;

QY 1 MHHHHHTAASDNFQSQGGGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAFASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 120
Db 61 NNGNGARVQVVGSAFASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 120
QY 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLIGIGI 172
Db 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLIGIGI 172
QY 173 PNQGGILYSSLEFEXKALEELAAAFPCDGLGSAADKYAGKNRNHVNFFQ-----ELAD 226
Db 173 PNQGGILYSSLEFEXKALEELAAAFPCDGLGSAADKYAGKNRNHVNFFQ-----ELAD 226
QY 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSSAGLMVAASPYVAMSVTAGQALTA 229
Db 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSSAGLMVAASPYVAMSVTAGQALTA 229
QY 227 LDRQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
Db 227 LDRQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
QY 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILITATLLGQNTPAIAVNEAEYGEWMA 287
Db 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILITATLLGQNTPAIAVNEAEYGEWMA 287
QY 276 LSAAFOAPFCAGAVAVVGGALAYLVVKTLINATQLLKLAKLAELVAAAIAD--IISDVAD 334
Db 276 LSAAFOAPFCAGAVAVVGGALAYLVVKTLINATQLLKLAKLAELVAAAIAD--IISDVAD 334
QY 288 QDAAMFPGYAAATATATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNVPO 347
Db 288 QDAAMFPGYAAATATATATATLLPFEAPEMTSAGGLLEQAAAVEASDTAAANQLMNVPO 347
QY 335 IIKGI-----LGEVWEFITNALNGLKELMDKLTGVTGLFSRGSNLESFFAG 382
Db 335 IIKGI-----LGEVWEFITNALNGLKELMDKLTGVTGLFSRGSNLESFFAG 382
QY 348 ALQOLAQPTGTTTSSKLGGLMKTVPSPRSP-----ISNMVSMANNHMTNSG 396
Db 348 ALQOLAQPTGTTTSSKLGGLMKTVPSPRSP-----ISNMVSMANNHMTNSG 396
QY 383 VPGLTGATSGLSQVTLGFGAAG-----LSASSGLAHADSLASSASLPALAG-----IGGG 432
Db 383 VPGLTGATSGLSQVTLGFGAAG-----LSASSGLAHADSLASSASLPALAG-----IGGG 432
QY 397 V-SMTNTLS--SMLKGFAPAAARQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRA 453
Db 397 V-SMTNTLS--SMLKGFAPAAARQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRA 453
QY 433 SGFGGLPSLAQVHAASRQALRPRADG-----PVGAAAEQVGGQSGLVSAQSGQMGG-PV 487
Db 433 SGFGGLPSLAQVHAASRQALRPRADG-----PVGAAAEQVGGQSGLVSAQSGQMGG-PV 487
QY 488 GMGCMHPSSGAS 499
Db 502 GQMGARAGGGLS 513

RESULT 8

US-10-098-732A-16
; Sequence 16, Application US/10098732A
; Publication No. US20030175294A1

```
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: Protein MTB72F (Ra12-TDH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
US-10-098-732A-16

Query Match      29.6%; Score 809; DB 12; Length 729;
Best Local Similarity 40.0%; Pred. No. 1.4e-54;
Matches 221; Conservative 69; Mismatches 170; Indels 92; Gaps 20;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVVGSAAPASLSIGTGDVITAVDGCAPINSATAMADALNHHFGDVISVTWQ 120
DB 61 NNGNGARVQVVGSAAPASLSIGTGDVITAVDGCAPINSATAMADALNHHFGDVISVTWQ 120

QY 121 TKSQGTGTGNVTLAEGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLIGIGI 172
DB 121 TKSQGTGTGNVTLAEGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLIGIGI 174

QY 173 PNOGGILYSSLEYFEKALEELAAAFPGDGLGSAADKYAGKNRNHNFFQ-----ELAD 226
DB 173 PNOGGILYSSLEYFEKALEELAAAFPGDGLGSAADKYAGKNRNHNHNFFQ-----ELAD 226

QY 175 -SVASDLFSAASAFQSVWVWGLTVG-----SWIGSSAGLMVAASAPYVAMWSVTAGQAEITA 229
DB 175 -SVASDLFSAASAFQSVWVWGLTVG-----SWIGSSAGLMVAASAPYVAMWSVTAGQAEITA 229

QY 227 LDRQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
DB 227 LDRQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275

QY 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILITNLGQNTPAIAYNEAEGEMWA 287
DB 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILITNLGQNTPAIAYNEAEGEMWA 287

QY 276 LSAFQAPFCAGAMAVVGGALAYLVVKTILINATQLKLLAKLAEVAAAIAD-IISDVAD 334
DB 276 LSAFQAPFCAGAMAVVGGALAYLVVKTILINATQLKLLAKLAEVAAAIAD-IISDVAD 334

QY 288 QDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNVVQ 347
DB 288 QDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNVVQ 347

QY 335 IIKGI-----LGEVWEFTNALNGKLKELWDKLTGWVTGLFSRGWSNLESFFAG 382
DB 335 IIKGI-----LGEVWEFTNALNGKLKELWDKLTGWVTGLFSRGWSNLESFFAG 382

QY 348 ALQQLAQPTQGTTPSSKGLGKLVSPHRSPI-----ISNMVSMANNHSMNTNSG 396
DB 348 ALQQLAQPTQGTTPSSKGLGKLVSPHRSPI-----ISNMVSMANNHSMNTNSG 396

QY 383 VPGLTGATSGLSQVTLFGAAG-----LSASSGLAHADSLASSASLPALAG-----ICGG 432
DB 383 VPGLTGATSGLSQVTLFGAAG-----LSASSGLAHADSLASSASLPALAG-----ICGG 432

QY 397 V-SMTNTLS--SMKGFAPAPAAQAQVTAQNGVYRAMSSLSGSLGSSGLGGVAAIIGRA 453
DB 397 V-SMTNTLS--SMKGFAPAPAAQAQVTAQNGVYRAMSSLSGSLGSSGLGGVAAIIGRA 453

QY 433 SGFGCLPSLAQVHAASTROALRPRADG-----PVGAALAEVQGGQQLVSAQSGQMGV-PV 487
DB 433 SGFGCLPSLAQVHAASTROALRPRADG-----PVGAALAEVQGGQQLVSAQSGQMGV-PV 487

QY 454 ASVSSL-SVPOWAAA--NOAVTPAARALPLTSLTSAARERFGQM-----LGGLPV 501
DB 454 ASVSSL-SVPOWAAA--NOAVTPAARALPLTSLTSAARERFGQM-----LGGLPV 501

QY 488 QMGGMHPSSGAS 499
DB 502 QMGGARAGGGLS 513
```

RESULT 9
US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US20030175294A1

```
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72FmutSA
; OTHER INFORMATION: (Ra12-TDH9-Ra35MutSA)
US-10-098-732A-18
```

```
Query Match      29.6%; Score 809; DB 12; Length 729;
Best Local Similarity 39.5%; Pred. No. 1.4e-54;
Matches 220; Conservative 66; Mismatches 169; Indels 102; Gaps 18;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVVGSAAPASLSIGTGDVITAVDGCAPINSATAMADALNHHFGDVISVTWQ 120
DB 61 NNGNGARVQVVGSAAPASLSIGTGDVITAVDGCAPINSATAMADALNHHFGDVISVTWQ 120

QY 121 TKSQGTGTGNVTLAEGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLIGIGI 172
DB 121 TKSQGTGTGNVTLAEGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLIGIGI 174

QY 173 PNOGGILYSSLEYFEKALEELAAAFPGDGLGSAADKYAGKNRNHNFFQ-----ELAD 226
DB 173 PNOGGILYSSLEYFEKALEELAAAFPGDGLGSAADKYAGKNRNHNHNFFQ-----ELAD 226

QY 175 -SVASDLFSAASAFQSVWVWGLTVG-----SWIGSSAGLMVAASAPYVAMWSVTAGQAEITA 229
DB 175 -SVASDLFSAASAFQSVWVWGLTVG-----SWIGSSAGLMVAASAPYVAMWSVTAGQAEITA 229

QY 227 LDRQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
DB 227 LDRQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275

QY 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILITNLGQNTPAIAYNEAEGEMWA 287
DB 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILITNLGQNTPAIAYNEAEGEMWA 287

QY 276 LSAFQAPFCAGAMAVVGGALAYLVVKTILINATQLKLLAKLAEVAAAIAD-IISDVAD 334
DB 276 LSAFQAPFCAGAMAVVGGALAYLVVKTILINATQLKLLAKLAEVAAAIAD-IISDVAD 334

QY 288 QDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNVVQ 347
DB 288 QDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNVVQ 347

QY 335 IIKGI-----LGEVWEFTNALNGKLKELWDKLTGWVTGLFSRGWSNLESFFAG 382
DB 335 IIKGI-----LGEVWEFTNALNGKLKELWDKLTGWVTGLFSRGWSNLESFFAG 382

QY 348 ALQQLAQPTQGTTPSSKGLGKLVSPHRSPI-----ISNMVSMANNHSMNTNSG 398
DB 348 ALQQLAQPTQGTTPSSKGLGKLVSPHRSPI-----ISNMVSMANNHSMNTNSG 398

QY 383 VPGLTGATSGLSQVTLFGAAG-----LSASSGLAHADSLASSASLPALAG-----ICGG 437
DB 383 VPGLTGATSGLSQVTLFGAAG-----LSASSGLAHADSLASSASLPALAG-----ICGG 437

QY 389 HNSMT--NGVSMNTVLTSSMLKGFAPAAQAQVTAQNGVYRAMSSLSGSLGSSGLGGV 446
DB 389 HNSMT--NGVSMNTVLTSSMLKGFAPAAQAQVTAQNGVYRAMSSLSGSLGSSGLGGV 446

QY 438 -----LPSLAQVHA--ASTROALRPRADG-----PVGAALAEVQGGQQLVSAQSGQMG 483
DB 438 -----LPSLAQVHA--ASTROALRPRADG-----PVGAALAEVQGGQQLVSAQSGQMG 483

QY 447 AANLGRAASVGSLSVPOWAAA--NOAVTPAARALPLTSLTSAARERFGQM-----L 496
DB 447 AANLGRAASVGSLSVPOWAAA--NOAVTPAARALPLTSLTSAARERFGQM-----L 496

QY 484 GG-PVGMGMHPSSGAS 499
DB 497 GGLPVGMGMGARAGGGLS 513
```

RESULT 10
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:

```

; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderman, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F-MAPS
; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Ra12-TbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
; US-10-098-732a-65

Query Match          29.6%; Score 809; DB 12; Length 930;
Best Local Similarity 39.5%; Pred. No. 1.9e-54;
Matches 220; Conservative 66; Mismatches 169; Indels 102; Gaps 18;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVHIGPTAFGLGVVD 60

Qy 61 NNGNGARVQRVGSAPASLSIGTGVITAVDGAPINSATAMADALNGHHPGDVI SVTWQ 120
Db 61 NNGNGARVQRVGSAPASLSIGTGVITAVDGAPINSATAMADALNGHHPGDVI SVTWQ 120

Qy 121 TKSGGTRTGNVTLAEGPPAEFLVPRGSM-----SRAFIIDP---TISADGLYDLLIGTI 172
Db 121 TKSGGTRTGNVTLAEGPPAEFLVPRGSM-----SRAFIIDP---TISADGLYDLLIGTI 172

Qy 173 PNOGGTLYSSLEVEEKALAEALAAFPDGLGSAADKYAGKRNHNVFFQ-----ELAD 226
Db 173 PNOGGTLYSSLEVEEKALAEALAAFPDGLGSAADKYAGKRNHNVFFQ-----ELAD 226

Qy 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSAGLMVAASPYVAMSVTAGQALTA 229
Db 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSAGLMVAASPYVAMSVTAGQALTA 229

Qy 227 LDRQLSLIHDOANAVQTTDRDILEGAKGLEFVRPVAVDL--TYIPVVG-----HA 275
Db 227 LDRQLSLIHDOANAVQTTDRDILEGAKGLEFVRPVAVDL--TYIPVVG-----HA 275

Qy 230 AQVRVAAAAYETAVGLTVPPVPI--AENRAELMILATNLGQNTPAIAVNEAEYGEWMA 287
Db 230 AQVRVAAAAYETAVGLTVPPVPI--AENRAELMILATNLGQNTPAIAVNEAEYGEWMA 287

Qy 276 LSAAFQAPFCAGAVVGVGALVVLVKTILNATQILKLLAKLAELVAAAIAD-IISDVAD 334
Db 276 LSAAFQAPFCAGAVVGVGALVVLVKTILNATQILKLLAKLAELVAAAIAD-IISDVAD 334

Qy 288 QDAAMFGYAAATATATATLLPPEEPMTSAGLLLEQAAVEEASDTAAANQOLMNNVPQ 347
Db 288 QDAAMFGYAAATATATATLLPPEEPMTSAGLLLEQAAVEEASDTAAANQOLMNNVPQ 347

Qy 335 IIKGI-----LGVWEFIITNALGKELWDLTGWTGLFSRGMNLSFFPAG 382
Db 335 IIKGI-----LGVWEFIITNALGKELWDLTGWTGLFSRGMNLSFFPAG 382

Qy 348 ALQOLAQPTGTPSSKLGGLWKTVPSPHSPI-----SNMVSMMANN 388
Db 348 ALQOLAQPTGTPSSKLGGLWKTVPSPHSPI-----SNMVSMMANN 388

Qy 383 VPGITGATSLGSLQVTLFGAAGLSASGLAHADSLASSASLPALAGIG---CGSGFGG--- 437
Db 383 VPGITGATSLGSLQVTLFGAAGLSASGLAHADSLASSASLPALAGIG---CGSGFGG--- 437

Qy 389 HMSGMT--NSGVSVMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSGLGGGV 446
Db 389 HMSGMT--NSGVSVMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSGLGGGV 446

Qy 438 -----LPSLAQVHA--ASTQALRPADG---PVGAAABQVGGGSLVSAQSQQM 483
Db 438 -----LPSLAQVHA--ASTQALRPADG---PVGAAABQVGGGSLVSAQSQQM 483

Qy 447 AANLGRAASVCSLSVPOWAAANQAVTPAARALPLISLSAERGFQM-----L 496
Db 447 AANLGRAASVCSLSVPOWAAANQAVTPAARALPLISLSAERGFQM-----L 496

Qy 484 GG-FVGMGGMHPSPSGAS 499
Db 484 GG-FVGMGGMHPSPSGAS 499

Qy 497 GGLPVGMGARAGGGLS 513
Db 497 GGLPVGMGARAGGGLS 513

RESULT 11
US-09-287-849-28
; Sequence 28, Application US/09287849
; Patent No. US20020009459A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040~
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; US-09-287-849-28

Query Match          27.0%; Score 739; DB 9; Length 231;
Best Local Similarity 100.0%; Pred. No. 8.1e-50;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 2 MHHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVHIGPTAFGLGVVD 61

Qy 61 NNGNGARVQRVGSAPASLSIGTGVITAVDGAPINSATAMADALNGHHPGDVI SVTWQ 120
Db 62 NNGNGARVQRVGSAPASLSIGTGVITAVDGAPINSATAMADALNGHHPGDVI SVTWQ 121

Qy 121 TKSGGTRTGNVTLAEGPPAEF 141
Db 122 TKSGGTRTGNVTLAEGPPAEF 142

RESULT 12
US-10-359-460-28
; Sequence 28, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
```


; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-28

Query Match
Best Local Similarity 27.0%; Score 739; DB 12; Length 231;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGVD 60
DB 2 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGVD 61
QY 61 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDI SVTWQ 120
DB 62 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDI SVTWQ 121
QY 121 TKSOGTGTGNVTLAEGPPAEF 141
DB 122 TKSOGTGTGNVTLAEGPPAEF 142

RESULT 13
US-09-841-132-337
; Sequence 337, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-841-132-337

Query Match
Best Local Similarity 26.9%; Score 737.5; DB 9; Length 585;
Matches 231; Conservative 40; Mismatches 174; Indels 151; Gaps 22;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGVD 60
DB 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGVD 56
QY 61 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDI SVTWQ 120
DB 57 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDI SVTWQ 116
QY 121 TKSOGTGTGNVTLAEGPPAEF
DB 117 TKSOGTGTGNVTLAEGPPAEF
QY 174 NOGQILYSSLE---YFEKA-----VPRGMSRAFIIDPTISAIDGLYDLGIG-IP 173
DB 117 TKSOGTGTGNVTLAEGPPAEFCRYPHWRPLDQVSESEPTSPD-----DVLGKG 168
QY 174 NOGQILYSSLE---YFEKA-----LEELAAAFPGDGLGSAADKAGKRNH 217
DB 174 VEGGINFQDLLEETRIKYNKAGTETKILTPSLQAQASAGNADAWASSPQSGS----- 228
QY 218 VNFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVA-----DLTY 268

DB 229 -----ATTVSDSGSDSSGSDTSETVPVTAKGGLYTDKNLSI 267
QY 269 IPVVGHALSAAFOAPFCAGAMAVVGGALAVLVKTLINATQLLKLAKLAELVAAAIADI 328
DB 268 TNITG-IIEIANNKATDVGGAYVGTL-----TCENSHR-LQFLKNSDCKGG----- 314
QY 329 ISDVADIIKIGLGEVWEFITNALNGKELWDLTGWVT--GLFSRG-----WSNLESF- 379
DB 315 -----GIYGEDNITLSN-LTG-KTLFQENTAKEBGGGLFIKTDKALTWTGLDSFC 363
QY 380 -----FAGVPGLTGATSGLSQVGTG--LFGAAGLSASSG 410
DB 364 LIINTSEKGGGAFVYKEISQTVTSVDETIPGIT-PVHGSETVITGNKSTGGGGVCTKR 422
QY 411 LAHADSLASASLPALAGIGGGGGLPFLSLAQVHAASRQALRRPADGVPVGAARQVGG 470
DB 423 LALSNLQISISGNSAAENGG-----AHTCPDSFPTAD-----TAQOPAA 463
QY 471 QSOLVSAQSGOGGVPVGMGHPSPSSGASKGTTTKYSEGAAAGTDEARAPVEAD 526
DB 464 ASAASTPKXS-----APVSTALSTPSSSTVSSLTLLAASSQASPATSNKETQDPNAD 515

RESULT 14
US-09-841-132-353
; Sequence 353, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 353
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-841-132-353

Query Match
Best Local Similarity 26.6%; Score 728; DB 9; Length 583;
Matches 228; Conservative 49; Mismatches 186; Indels 140; Gaps 21;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGVD 60
DB 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGVD 56
QY 61 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDI SVTWQ 120
DB 57 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDI SVTWQ 116
QY 121 TKSOGTGTGNVTLAEGPPAEF-----LVPRGMSRAFIIDPTISAIDGLYDLGIG 171
DB 117 TKSOGTGTGNVTLAEGPPAEFCRYPHWRPLDQVSESEPTSPD-----DVLGKG 168
QY 172 IPNOGQILYSSLEYFEKALEELAAAFPGDGLGSAADKIYAG-----KNRNHVNFFQ 222
DB 169 -----GGI-----YTEKSLTITGITIDFVSNIAIDSGAGVFTKENLSCTNLSLQFLK 218
QY 223 ELADLD-----RQLISLIHQANAVQTRDIL--EGAKKG-----LEFVR 260
DB 219 NSAGOHGGGAVYVQTMSTVNTTSESITTPPLVGEVIFSENTAKHGCGGICTNKLKSLNKL 278
QY 261 PVAV-----DLTYIPVVGHALSAAFOAPFCAGAMAVVGGALAVLVKTLIN 306
DB 279 TVTLTKNSAKESGGAIPTDLASIPITDTPESSTPSSSPASTPPEVVA----- 327

Qy	307	ATQLKLIKLAELVAAAIIDISDVAD-----ITKGILGEVWFIFITNALNG	355
Db	328	---KINFFASTAEAPAPSLTAEBSQDQDTSTSDIDVSLINLVAIQNTSAKKG	385
Qy	354	LKELWDKLTGWVTGLFSR--GWSNLEFFAGVPEGLTGATS-----GLSQVTGL-PCAAGL	405
Db	386	-----GALYKKAKLKRINLLE-----LSGNSSQDVGGGLCLTSEVFEPAIG-	427
Qy	406	SASSGIAIADSLASSASL-----PALAGIGGSGFGGLPSLQVHAASITQALRPADCP	460
Db	428	---SLLSHYNSAKEGGVHISKTVTLGNLKSFTTFADNTVKAIVESTPEAPEIIPVEGE	484
Qy	461	VGAAAE---QVGGOSQLVSAQSGCMGPPVCMGGMHPPSSCAKSGTKTTTKYSEGAAAGTE	516
Db	485	ESTATENPENSNTGESSANTNLEGSGQDTADTGIGVAVNESQDTSIDICNAESGELQDSTQ	544
Qy	517	DAE 519	
Db	545	SNE 547	

RESULT 15

US-09-841-132-333
; Sequence 333, Application US/09841132
; Patent No. US20020061848A1

; GENERAL INFORMATION:

APPLICANT: Bhatia, Ajay

APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Probst, Peter

REF. CRYST. 1955/1000
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

1. TITLE OF INVENTION: CORRELATIONS AND METHODS FOR PREVENTION OF INFECTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

FILE REFERENCE: 210121.469C8

FILE REFERENCE: Z19121-10200
: CURRENT APPLICATION NUMBER: US/09/841,132

; CURRENT APPLICATION NUMBER: US/00-04-23
: CURRENT FILING DATE: 2001-04-23

;; CURRENT FILING DATE: 2001
: NUMBER OF SEQ' ID NOS: 599

```

; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0

```

; SOFTWARE: Ed
; SEQ ID NO 333

```

; SEQ ID NO 333
: LENGTH: 518

```

```
; LENGTH: 518
; TYPE: DPT
```

TYPE: PRT
ORGANISM: Chlamydia trachomatis

; ORGANISM: CHAI
US-08-841-132-333

Query Match	26.4%	Score 722.5;	DB 9;	Length 518;
Best Local Similarity	37.3%	Pred. NO. 4.8e-48;		
Matches	209;	Conservative	44;	Mismatches 132;
				Indels 175;
				Gaps 21;

Qy	1	MHHHHHTTAASNFOLSGGGGFAIPITQOAMAIAGQIRSGGSPFVHI	GPTAFPLGLGVDD	60	
Db	1	MHHHHHTTAASNFOLSGGGGFAIPITQOAMAIAGQIK----	LPTVHI	GPTAFPLGLGVDD	56
Qy	61	NNNGARVORVGSAPASLSIGTGDVITAVDGGAPINSATAMADALNGHHPGDVISVTWQ	120		
Db	57	NNNGARVORVGSAPASLSIGTGDVITAVDGGAPINSATAMADALNGHHPGDVISVTWQ	116		
Qy	121	TKSGGTRTGNVLABGPPAEF-----	IVPRGSMRAP-IDPTISAGIDLXDLGI	170	
Db	117	TKSGGTRTGNVLABGPPAEFCRYFSHWRPLMKWLSATAVFAAFLPSVS	-----GF	167	
Qy	171	GIPNOGGIYSSLE-----	YFEKALEELAAAFPGDGWLG-----	SADKYAGKNHHV	218
Db	168	CFPEPKELNFSRVETSSSTFTTETIGEGAGYIVSGNASFTKFTIPTD	TTTTPTTNSSS	227	
Qy	219	NFQOEALDLDRQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDLTYIPVGH	HALSA	278	
Db	228	SSSGETAS-----	VSEDSSTTTTTD-----	PKG-----	251
Qy	279	AFQAPFCAGAMAVVGGALAYLVVTKLINATOLLLAKLAELVAAAIADIIS	VADIIGK	338	
Db	252	-----GGA-----	-----FYNAAH-----	-----SG	261
Qy	339	ILGEWVEFTNALNGKELWDLKTGWVTGLFSRGSWNLSEFPFAGVPOL	TGATSEL	SVQVG	398
Db	262	VLGFMTRSCTEGSLTISEI--RMTGEGAIISQO-----	ELLFTD	LTSLT--QNNLSQ	314

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 11.2444 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-64

Perfect score: 2737

Sequence: 1 MHHHHTAASDNQLSQGG.....RAPVEADAGGQKVLVRNV 539

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pir_76.*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1963	71.7	392	2 A70957	hypothetical prote
2	1263	46.1	394	2 S72814	hypothetical prote
3	676	24.7	355	2 F70993	probable serine pr
4	474.5	17.3	361	2 S47170	hypothetical prote
5	460.5	16.8	354	2 A87242	probable secreted
6	385.5	14.1	402	2 E70656	hypothetical prote
7	199.5	7.3	464	2 C70821	probable serine pr
8	188.5	6.9	382	2 H86930	probable serine pr
9	188.5	6.9	452	2 T45448	probable serine pr
10	173	6.3	837	2 H72802	minor tail subunit
11	169	6.2	496	2 H70839	hypothetical glyci
12	168	6.1	1417	2 A83080	hypothetical glyci
13	167	6.1	916	2 T03323	gene 116 protein -
14	166.5	6.1	1018	2 H83135	probable adhesin p
15	165	6.0	864	1 EART	elastin precursor
16	164	6.0	837	2 S30971	gene 26 protein -
17	159.5	5.8	860	1 FAMS	elastin precursor
18	159	5.8	1467	2 A75564	conserved hypothet
19	154	5.6	1329	2 E70917	hypothetical glyci
20	152.5	5.6	914	2 H70987	hypothetical glyci
21	152	5.6	409	2 A70647	probable PPE prote
22	151.5	5.5	940	2 F84089	phage-related prot
23	151	5.5	588	2 F70971	hypothetical glyci
24	150	5.5	439	2 D70954	hypothetical glyci
25	150	5.5	741	2 G70917	hypothetical glyci
26	150	5.5	801	2 F70824	hypothetical glyci
27	150	5.5	1306	2 T09324	hypothetical glyci
28	150	5.5	1517	2 T13329	hypothetical prote
29	149.5	5.5	749	2 A70812	hypothetical glyci

30	149	5.4	783	2 E70824	hypothetical glyci
31	149	5.4	2174	2 E95965	hypothetical glyci
32	148.5	5.4	1510	2 T13234	probable minor tail
33	148	5.4	1650	2 A70869	hypothetical glyci
34	147.5	5.4	539	2 A70899	probable PPE prote
35	147.5	5.4	684	2 T36771	probable integral
36	147	5.4	732	1 EAHU	elastin precursor
37	147	5.4	960	1 A39651	discs-large tumor
38	147	5.4	1381	2 E70806	hypothetical glyci
39	146.5	5.4	515	2 H70563	hypothetical glyci
40	146	5.3	805	2 T25795	hypothetical prote
41	146	5.3	957	2 D70835	hypothetical glyci
42	145.5	5.3	667	2 A70893	hypothetical glyci
43	145.5	5.3	694	2 F70868	hypothetical glyci
44	145.5	5.3	882	2 B70812	hypothetical glyci
45	143.5	5.2	837	2 E70835	hypothetical glyci

ALIGNMENTS

RESULT 1

A70957

hypothetical protein Rv3616c - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: A70957

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

R/Connor, R.; Davies, R.; Devlin, K.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

R/Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: A70957

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-392 <COL>

A/Cross-references: GB:Z95557; GB:AL123456; NID:G3242276; PIDN:CAB08950.1; PID:e316833; I

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: RV3616c

Query Match 71.7%; Score 1963; DB 2; Length 392;

Best Local Similarity 99.7%; Pred. No. 2.4e-108;

Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTTISAIDGLYDGLLIGIPNOGGILYSSLEYFEKALEBELAAAFPGDGLGSA 207

DB 1 MSRAFIIDPTTISAIDGLYDGLLIGIPNOGGILYSSLEYFEKALEBELAAAFPGDGLGSA 60

QY 208 DYAGKNNHNHVFPELADRLQLSLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 267

DB 61 DYAGKNNHNHVFPELADRLQLSLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

QY 268 YTPVVGHALSAAFOAPFCAGAMAVGGALVIVKLTINATQLLAKLAELVAAATAD 327

DB 121 YTPVVGHALSAAFOAPFCAGAMAVGGALVIVKLTINATQLLAKLAELVAAATAD 180

QY 328 IISDVADIKGIIGVWEFFITNALNGLKELDKLTGWVTGLFSRGSNLSFFAGVPGIT 387

DB 181 IISDVADIKGIIGVWEFFITNALNGLKELDKLTGWVTGLFSRGSNLSFFAGVPGIT 240

QY 388 GATSGLSQVTGLFGAGLSASSGLAHADSLASSLPALAGIGGSGFGGLPSLAQVHAA 447

DB 241 GATSGLSQVTGLFGAGLSASSGLAHADSLASSLPALAGIGGSGFGGLPSLAQVHAA 300

QY 448 STRQALRPADGPVGAAGAEQVGGSQLVSAQSGQMGPVGVGMHPSGASKGTTTKY 507

DB 301 STRQALRPADGPVGAAGAEQVGGSQLVSAQSGQMGPVGVGMHPSGASKGTTTKY 360

QY 508 SEGAACTEDERAPVEADAGGQKVLVRNV 539

Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 2

S72814

hypothetical protein B1620.C2_213 - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001

C:Accession: S72814

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid B1620.

A:Reference number: S72584

A:Accession: S72814

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <SMI>

A:Cross-references: EMBL:U00015; NID:G466931; PIDN:AAC43223.1; PID:G466935

Query Match 46.1%; Score 1263; DB 2; Length 394;

Best Local Similarity 62.9%; Pred. No. 3.6e-67;

Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;

QY 148 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALELEAAAFPGDGLGSA 207

DB 1 MSGAFIIDPTLKAIEAWHALLG:GVPNDGGVLYSSLSPEKALEHLAAAFPGDGLGSA 60

QY 208 DKYAGQNRHNFQELADLDQLISLHDQANAVOTTDELEGAKKGLFVRPVAVDLT 267

DB 61 DKYAGQNRKRDVDFQELADLDKELIELIHQANSVOTTRIGLDGAKKALLFVRPVAIDL 120

QY 268 YIPVWGHALSAAPQAPFCACAMAVGALAYLVVKTLINATQLKLLAKLAEVAAAIAD 327

DB 121 YIPVGSVNSASIQACAAAMAAVSGGLAYLVVQTAIHTAKFVALLARLALLASAVAD 180

QY 328 IISDVADIINGLGEVWEFETNALNGKLMDKLTGWVTLFSGWNSNLESFFAGVPGIT 387

DB 181 VSDGVAIIKGIVDHLWHFETAGALTGLKDIKVEIHHWFGLFSHWSRLHSPFGIPGLS 240

QY 388 GATSGLSQVTLGFGAAGLASSGLAHADSLASSPALAGIGGGSGFGLPSLAQVHAA 447

DB 241 GATSGLSQVTLGFGVGLAGSSGLSLLSTENLPSLAGVAGLGLSLQLAQHLAA 300

QY 448 STRQALRPADGPGVGAAGQVQSQSLVSAQSGQGGVPGVGMHPSSGASK--GTTTK 505

DB 301 STRQTRSQAGVSAELSTEQFGQEPVSAQSGQGGGQGGMGMPASTKSKDKERKK 360

QY 506 KYSEGAAGCTDAERAPVEADAGGQKVLVRNV 539

DB 361 KYSEGAAGTDAERAPVEQSGGKRAUAQHVV 394

RESULT 3

F70983

probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C:Accession: F70983

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70983

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-355 <COL>

A:Cross-references: GB:Z96071; GB:AL123456; NID:G3242254; PIDN:CAB09453.1; PID:G2181967

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: pepA

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps

Query Match 24.7%; Score 676; DB 2; Length 355;

Best Local Similarity 99.2%; Pred. No. 1e-32; 0; Indels 0; Gaps 0;

Matches 132; Conservative 1; Mismatches 0;

QY 7 HTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGA 66

DB 223 NTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGA 282

QY 67 RVQVVGSAPASLGLSTGCVITAVDGPINSATADALNGHHHPGDVLSVTWTKSGGT 126

DB 283 RVQVVGSAPASLGLSTGCVITAVDGPINSATADALNGHHHPGDVLSVTWTKSGGT 342

QY 127 RTGNVTLAEGPPA 139

DB 343 RTGNVTLAEGPPA 355

RESULT 4

S47170

hypothetical protein 34K - Mycobacterium paratuberculosis

C:Species: Mycobacterium paratuberculosis

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Dec-2002

C:Accession: S47170

R:Camerton, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.

submitted to the EMBL Data Library, June 1993

A:Description: Isolation and characterisation of a 34kDa protein of Mycobacterium paratub

A:Reference number: S47170

A:Accession: S47170

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-361 <CAM>

A:Cross-references: EMBL:Z23092; NID:G505550; PIDN:CAA80638.1; PID:G505551

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps

Query Match 17.3%; Score 474.5; DB 2; Length 361;

Best Local Similarity 69.7%; Pred. No. 7.6e-21;

Matches 92; Conservative 17; Mismatches 22; Indels 1; Gaps 1;

QY 8 TAAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 67

DB 231 TAATDSYKMS-GGQGFAPIGRAVAVANQIRSGAGSNVTHIGPTAFGLGVVDNNGGAR 289

QY 68 VQVVGSAPASLGLSTGCVITAVDGPINSATADALNGHHHPGDVLSVTWTKSGGT 127

DB 290 VQVVNTGPAAGIAPGDVITVTPINGATSMTELVVPHHPGDTIAVHFRSVDGGER 349

QY 128 TGNVTLAEGPPA 139

DB 350 TANITLAEGLPPA 361

RESULT 5

A87242

probable secreted serine proteinase [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002

C:Accession: A87242

R:Coile, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, R.; Davies, R.M.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squ

A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: A87242

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-354 <STO>

A:Cross-references: GB:AL450380; NID:G13093863; PIDN:CAC32191.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML2659

Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 2

S72814

hypothetical protein B1620.C2_213 - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001

C:Accession: S72814

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid B1620.

A:Reference number: S72584

A:Accession: S72814

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <SMI>

A:Cross-references: EMBL:U00015; NID:G466931; PIDN:AAC43223.1; PID:G466935

Query Match 46.1%; Score 1263; DB 2; Length 394;

Best Local Similarity 62.9%; Pred. No. 3.6e-67;

Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;

QY 148 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALELEAAAFPGDGLGSA 207

DB 1 MSGAFIIDPTLKAIEAWHALLG:GVPNDGGVLYSSLSPEKALEHLAAAFPGDGLGSA 60

QY 208 DKYAGQNRHNFQELADLDQLISLHDQANAVOTTDELEGAKKGLFVRPVAVDLT 267

DB 61 DKYAGQNRKRDVDFQELADLDKELIELIHQANSVOTTRIGLDGAKKALLFVRPVAIDL 120

QY 268 YIPVWGHALSAAPQAPFCACAMAVGALAYLVVKTLINATQLKLLAKLAEVAAAIAD 327

DB 121 YIPVGSVNSASIQACAAAMAAVSGGLAYLVVQTAIHTAKFVALLARLALLASAVAD 180

QY 328 IISDVADIINGLGEVWEFETNALNGKLMDKLTGWVTLFSGWNSNLESFFAGVPGIT 387

DB 181 VSDGVAIIKGIVDHLWHFETAGALTGLKDIKVEIHHWFGLFSHWSRLHSPFGIPGLS 240

QY 388 GATSGLSQVTLGFGAAGLASSGLAHADSLASSPALAGIGGGSGFGLPSLAQVHAA 447

DB 241 GATSGLSQVTLGFGVGLAGSSGLSLLSTENLPSLAGVAGLGLSLQLAQHLAA 300

QY 448 STRQALRPADGPGVGAAGQVQSQSLVSAQSGQGGVPGVGMHPSSGASK--GTTTK 505

DB 301 STRQTRSQAGVSAELSTEQFGQEPVSAQSGQGGGQGGMGMPASTKSKDKERKK 360

QY 506 KYSEGAAGCTDAERAPVEADAGGQKVLVRNV 539

DB 361 KYSEGAAGTDAERAPVEQSGGKRAUAQHVV 394

RESULT 3

F70983

probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C:Accession: F70983

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70983

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-355 <COL>

A:Cross-references: GB:Z96071; GB:AL123456; NID:G3242254; PIDN:CAB09453.1; PID:G2181967

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: pepA

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps

Query Match 46.1%; Score 1263; DB 2; Length 394;

Best Local Similarity 62.9%; Pred. No. 3.6e-67;

Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;

QY 148 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALELEAAAFPGDGLGSA 207

DB 1 MSGAFIIDPTLKAIEAWHALLG:GVPNDGGVLYSSLSPEKALEHLAAAFPGDGLGSA 60

QY 208 DKYAGQNRHNFQELADLDQLISLHDQANAVOTTDELEGAKKGLFVRPVAVDLT 267

DB 61 DKYAGQNRKRDVDFQELADLDKELIELIHQANSVOTTRIGLDGAKKALLFVRPVAIDL 120

QY 268 YIPVWGHALSAAPQAPFCACAMAVGALAYLVVKTLINATQLKLLAKLAEVAAAIAD 327

DB 121 YIPVGSVNSASIQACAAAMAAVSGGLAYLVVQTAIHTAKFVALLARLALLASAVAD 180

QY 328 IISDVADIINGLGEVWEFETNALNGKLMDKLTGWVTLFSGWNSNLESFFAGVPGIT 387

DB 181 VSDGVAIIKGIVDHLWHFETAGALTGLKDIKVEIHHWFGLFSHWSRLHSPFGIPGLS 240

QY 388 GATSGLSQVTLGFGAAGLASSGLAHADSLASSPALAGIGGGSGFGLPSLAQVHAA 447

DB 241 GATSGLSQVTLGFGVGLAGSSGLSLLSTENLPSLAGVAGLGLSLQLAQHLAA 300

QY 448 STRQALRPADGPGVGAAGQVQSQSLVSAQSGQGGVPGVGMHPSSGASK--GTTTK 505

DB 301 STRQTRSQAGVSAELSTEQFGQEPVSAQSGQGGGQGGMGMPASTKSKDKERKK 360

QY 506 KYSEGAAGCTDAERAPVEADAGGQKVLVRNV 539

DB 361 KYSEGAAGTDAERAPVEQSGGKRAUAQHVV 394

RESULT 3

F70983

probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C:Accession: F70983

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70983

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-355 <COL>

A:Cross-references: GB:Z96071; GB:AL123456; NID:G3242254; PIDN:CAB09453.1; PID:G2181967

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: pepA

Qy	262	VAVDLTVIPVGHALS---	AAQAPFCAGAMAVVG-----	293	
Db	11	VAAAATHLAGIGSALS	TANAAAAAP--TTALS	VAGADESVLIAALFEAYAEQVQALS	68
Qy	294	-----	GALAYLVVKTLINATOLLK	LAKLAEVAAAIADIISDVADII---	336
Db	69	ALAFHDQFQALNNGAVCY	AAAAET--ANATPL-----	QALQTVOQNVLTVVNAP	115
Qy	337	KGILGEWBEFITNALNGL	KELDKLTGCVWG---LFSRGWNLSEFF	PAGVFGLTGATS	391
Db	116	TQALLGR--PIINGANGL	FN-----TGQDGGPGGLLF	GNGGN-----GGSG	155
Qy	392	GLSQVTGLFGAAGL--	SASGLAHADSLASSASLP	A---LAGIGGSGFGGLPSLAQV	444
Db	156	GVDQAGNGGAAAGLIGN	GSGGVPICISAGGAGGAGGL	LFNGPGPGAGGIGITGD-	214
Qy	445	HAASRQALPRADPGVCA	AEVGGOSQLVSAGSOGCMGPV	GMGMHPSGASKGTIT	504
Db	215	-----GGPGGAGNATG	-----LFGSGGTGGMGV	GMGGMVGNNGVAGNGGTA	257
Qy	505	KKYSEGAAAGTDAERAP	VEADAGG	530	
Db	258	GLFGHGAGGAGGIGTS	ADGLGGGGG	283	

RESULT 12

A83080

hypothetical protein PA4541 [imported] - *Pseudomonas aeruginosa* (strain PAO1)

C:Species: *Pseudomonas aeruginosa*

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: A83080

R:Scover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen

A:Reference number: A83950; MUID:20437337; PMID:10984043

A:Accession: A83080

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1417 <STO>

A:Cross-references: GB:AE004867; GB:AE004091; NID:g9950769; PIDN:AA07929.1; GSPDB:GN001

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: PA4541

```

Query Match      6.1%; Score 168; DB 2; Length 1417;
Best Local Similarity 22.1%; Pred. No. 0.042;
Matches 139; Conservative 77; Mismatches 206; Indels 206; Gaps 31;

QY      19  GGCGFAIPGQMAAIG-----QIRSG 41
      |||
Db       782  GGNSSFLSIGNAFGGTFPCLGNTDNLAVYGTGAYGSLFSVNRGTLNMLERISADG 841
QY      42  GSTT---VHIGTFAFLGLGVDDN-NGNGARVQRVWGSAPAASLGISTGVVITAVDGAPIN 97
      |||
Db       842  AQATHYNVQVGSAAVNLGSRIDNVNASDIRI-----AAASKLNSLGG-LVALNLGSD 693
QY      98  SATAMADALNGHHPGDVISVTWQTKGGIRTGNTLAEG-----PBAFLVP----- 144
      |||
Db       894  NASASGTLVGNRH-----TYAL--GLLAENISTARGVASISNRRADFALSGQLKHA 944
QY      145  -----FGMSRAFIIDPTISAIDGLVDLLGIGIPNQGGL-YSSLEYFEKALE 191
      |||
Db       945  SHYGAGGLVGRNRGGIIRS-----SGSQGTLSLSGHGM-NLGLLVGYSS-----AGGLA 992
QY      192  ELAAA--FPGDGWLGSAADKYACKNRHNHNFQELADLRQLLSLHDCANAV--QTTRD 247
      |||
Db       993  DVASVDVSGNGRG-----LYGGLIGLVN-----SGIAHATASGKVRGTDAE 1036
QY      248  ILGSAKKGLEFVRPAVDLTYIPVWGHJASAAFPFCAGAMAV-----VGGALAY----- 298
      |||

```

```
D$      1037 ALGG-----LIGRNUNAAINNNAHGDVSLQAGRYLGGLIGHNQAG   1077
QY      299 -----LVYKTIINATQLKLAKLAELVAATAIISVDADIKGIL-----    340
D$      1078 NLNVSTSGNLSGGSLAQGGLIGNANASLVNASAKGNVATRCAEAVGGLGENLYGSV   1137
QY      341 -----GEWWEFITNALNGLKELMDKLTGWVTCLFSRGHSNLSEFFACVPGLTCATSGLS   394
D$      1138 INGSASGEVTTDGSKTLLGL--IGSNLGGNHNSLNKSGMWYN-----AGAN---S   1181
QY      395 QVTGLFGAAG-SASSGLAHADSASSASLPALAG--IGGSGFPGGLPSLAQHHAASTROA   452
D$      1182 DVGGJIG----HNREGGSHS-TLASGNTVGKGSRVGLGVGYNDAASSLTNVSASGNVSA   1235
QY      453 LRPRADGPVGAARAEQGCGSQLVASQG-----SQMGGPVMGM-----GMHPSSGASKG   501
D$      1236 SG$RAIG- -GIUGDLRSLMLASHGIVYDKITSHNLGGLVGRGENTSIRSAKASGAVSG   1293
QY      502 TTKKYSEGAAAGTEDAERPVEADAGG   529
D$      1294 CAGIR-AGGLVGSLEGWOAILICAGAGG   1320
```

```

RESULT 13
T03323
gene l16 protein - Lactococcus phage b11170
C:Species: Lactococcus phage b11170
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 05-Nov-1999
C:Accession: T03323
R:Crutz-Le Coq, A.M.; Cesselin, B.; Commissaire, J.; Anba, J.; Kyriakidis, S.; Chopin, M.
submitted to the EMBL Data Library, June 1997
A:Description: Sequence and organization of the lactococcal isometric b11170 phage genome
A:Reference number: Z14903
A:Accession: T03323
A:Status: preliminary; translated from GB/EMBL/DBEJ
A:Molecule type: DNA
A:Residues: 1-916 <CRU>
A:Cross-references: EMBL:AF009630; NID:g3282260; PIDN:AAC27195.1; PID:g3282276
C:Genetics:
A:Gene: l16

```

Query Match 6.1%; Score 167; DB 2; Length 916;
Best Local Similarity 21.8%; Pred. No. 0.029;
Matches 119; Conservative 70; Mismatches 189; Indels 168; Gaps 25;

```

51  TAFI:GLG-----VVDNN-----GNGARVORVW---GSAPAA:SLGISTGDVITAV 91
    :|||:
229  TTFI:GLGSAKTAVSKTEALYKANQAFG:GTGNLKGWQAYGQMSAAG-----KV 278
    :|||:
92  DGAP:INSATAVADALNGHHPGDVISVTWQTKSGGT-----RTGNVTL-----AEG 136
    :|||:
279  TAEN:INQUTDNNTALGASLKUTVMQNPQLKOYGSFNEAVSAGAVSMDMLDKAMQKAADG 338
    :|||:
137  PPAE:FLVPRGS-----MSRAFIIDPTISA-----IDGLYDLLIGIPNQGGILYSLEY 185
    :|||:
339  SSSA:TKI:RTDTSWGFNEDLSQALI--PTLEALT:PVINALIDKDDMGKAGKAIENVVKY 396
    :|||:
186  PE---KALEE-----LAAAPG-DGWLGSAAK:YAG-----KN-----RNHV 219
    :|||:
397  FQDL:FKOLQOQNGAITQFSAIWNLKSAFGSVIGI:GNLK:SPAGVDESTSKNSTSVENVA 456
    :|||:
219  NFFO:ELADLDRQL:SLIHDQANAVOTTRDIL:EGAKKGLFEFVPVAVDL-----T 267
    :|||:
457  NTIS:SLANKPADIITK:IAIDFTGKISEKSAMDAI:KVALVALAGAFVAMKVINGI:IKAYET 516
    :|||:
268  YIPV:V---GHALSA:FOAPFCAGAMAVVGGALAYL:VVKTLINATQL---LKLAKLAELVA 322
    :|||:
517  YNKI:VEAGTIIQGAFA:IMAVNPPVLLG:IAAAVAVAGVYFFETQ:ETGKKAWASPFVDFLK 576
    :|||:
323  AA-----TADI:SDVADI:IKGI-----LGEVWEFI-----347
    :|||:
577  SAMD:GIVSFFSGIGQW:FA:DIWNGAVDGA:KIQWGLVDWFI:GIVQGIQNTWNGITTFPTTL 636
    :|||:

```

QY 348 -TNALNGLKELDKLTGWVTGLF-----SRGWSNLESPFAGVPGITG 388
Db 637 WTVVIGIGSVGGVGTGFGSIFDAVSVVSTVFSFSAIGGFASSAMNVLVSWSAVAGFFG 696
QY 389 ----ATSGLSQVTLGFLGAGLASSGLAHADSLASAS-----LPALAGIGG--SFGF 436
Db 697 GIFNAVSGV--VSSVFSFSAIGGFASSAMNVLVSWSAVAGFFG 754
QY 437 GLPSLA 442
Db 755 GFASNA 760

RESULT 14
H83135
Probable adhesin PA4082 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83135
R:Stover, C.K.; Phan, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1018 <STO>
A:Cross-references: GB:AE004824; GB:AE004091; NID:g9950277; PIDN:AAG07469.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4082

Query Match 6.1%; Score 166.5; DB 2; Length 1018;
Best Local Similarity 23.7%; Pred. No. 0.035;
Matches 149; Conservative 75; Mismatches 259; Indels 145; Gaps 28;

QY 5 HHTAASDNFQLSQGGGFAIPICQAMATAGQIRSGGSPTHVIGTFLGLGVVDN--- 61
Db 174 YRFTGTPSTNGLNHGGAITAAEGGSIALGGAQVDNRG---TV-LAQMGVGVGAGSDTLT 229
QY 62 NNGARVORVVGAPAAASLGISTGDIVTAVDGAPINSATAMADALNG---HHPGDVISVT 118
Db 230 NFGNKLDIRVDAGVANALASNGGLKA-DGRVLMARTANALLNTVVNSQGA--- 284
QY 119 WQTKSGGTRTGNVTLAEGPPAEFLVPRGMSRAFIIDPTISAIDGLYLLGIGIPNQGI 178
Db 285 -EARSLRGKNGRIVLDGGPGKVMV--GGALS-----ANALNG-----FQHGQT 325
QY 179 LYSLEVFKEALAEALAAFPDGLGSG---AADKY-----AGKRNHNVFFQELADL 227
Db 326 VEVRGQAVEALGTQVNTLASNGLNGTWKTAADKIDVRSAVSDGTVTVADTLRNLAST 385
QY 228 DROLISLIHD-----QANAVQTRTIDLECAKGLGFVRPVAVDLRY 268
Db 386 NIELVSTKGBDLDDGSVNWASGNRLGLGSAADTLNGLNASGAKAGLEKAEGAIDIND 445
QY 269 IPVVGHALSA-----AFQAPFCAGAMAV-VGGALAVLVYKTLINATQLLKLA 315
Db 446 KIVLGGAGSALAMDAGSGHRVNTASVSLAGANATVYSGGYTYVONLAQIQAINKLD 505
QY 316 KLAELVAAAIADIISDVADIKIGILEWFEFFITNALNGLKELDKL-----TGWVTGLFGR 371
Db 506 GLYVLGNGILGG--SYCYTALQSIGGPAGVF-SGTLDLGLNSIGNLSISNTGPNVGLFAR 562
QY 372 G----WSNL-----BSFPAGVPLGTGATSGLSQVTLGFGAAGLASSGLAHADSLAS 420
Db 563 SSGTSLNKLNNLRVSDNTYSGSPSSIGALVNGS-----GRIANVSAGSVSVVGRSLRS 617
QY 421 ASLPALAG-----IGGSGFGGLPLSLAQVHAASTROALRPADGPGVGAAGVGGQSOL 474
Db 618 NALGGLVGRNISGQIANASVSGGVGTG-----YANAST-----AVGGLVGENFTTAWGPEAV 667

QY 475 -----VSAQGSQ-----GMGGPVGMG--GMHPSGASKGTTHKKYSEGAAAG----- 514
Db 668 IENAHSNVHVAQSGSTERNSLGGVGLVGLNAKGMIRASGSGQKRVET--YRPLNVLGGLVG 725
QY 515 -----TEDAERAPVEADAGGQKQVLV 535
Db 726 YNMFGRVSDSSASGQVEAGGAGNTGGLV 753

RESULT 15
EART
elastin precursor - rat
N:Alternate names: tropoelastin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jan-1991 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
C:Accession: A36106; A30878; A36523; S02173; I54172; I68505
R:Pierce, R.A.; Deak, S.B.; Stolle, C.A.; Boyd, C.D.
Biochemistry 29, 9677-9683, 1990
A:Title: Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.
A:Reference number: A36106; MUID:91104868; PMID:1702999
A:Accession: A36106
A:Molecule type: mRNA
A:Residues: 1-864 <PIE>
A:Cross-references: GB:M60647; GB:J05292; NID:G207444; PIDN:AAA42269.1; PID:G207445
R:Deak, S.B.; Pierce, R.A.; Belsky, S.A.; Riley, D.J.; Boyd, C.D.
J. Biol. Chem. 265, 13504-13507, 1988
A:Title: Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.
A:Reference number: A30878; MUID:88330868; PMID:2971041
A:Accession: A30878
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 781-864 <DEA>
A:Cross-references: GB:J04035; NID:G207442; PIDN:AAA42268.1; PID:G207443
R:Franzblau, C.; Pratt, C.A.; Paris, B.; Colaninno, N.M.; Offner, S.D.; Mogayzel Jr., P.
J. Biol. Chem. 264, 15115-15119, 1989
A:Title: Role of tropoelastin fragmentation in elastogenesis in rat smooth muscle cells.
A:Reference number: A36523; MUID:89359327; PMID:2768256
A:Accession: A36523
A:Molecule type: protein
A:Residues: 22-31 <FRA>
R:Rich, C.B.; Foster, J.A.
Arch. Biochem. Biophys. 268, 551-558, 1989
A:Title: Characterization of rat heart tropoelastin.
A:Reference number: S02173; MUID:89117149; PMID:2913947
A:Accession: S02173
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1P, 369-545, 548-764, 770-864 <RIC>
A:Experimental source: heart
R:Pierce, R.A.; Alatawi, A.; Deak, S.B.; Boyd, C.D.
Genomics 12, 651-658, 1992
A:Title: Elements of the rat tropoelastin gene associated with alternative splicing.
A:Reference number: I54172; MUID:92241859; PMID:1572637
A:Accession: I54172
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 264-533 <RES>
A:Cross-references: GB:M86372; NID:G207455; PIDN:AAA42271.1; PID:G554527
A:Accession: I68505
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 558-864 <RE2>
A:Cross-references: GB:M86376; NID:G207459; PIDN:AAA42272.1; PID:G207462
C:Genetics:
A:Introns: 277/1; 292/1; 308/1; 339/1; 359/1; 419/1; 437/1; 467/1; 484/1; 601/1; 621/1; 6
A:Note: the list of introns may be incomplete
C:Superfamily: elastin
C:Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-864/Product: elastin #status predicted <MA>
F:854-859/Disulfide bonds: #status predicted


```

Query Match      6.0%; Score 165; DB 1; Length 864;
Best Local Similarity 23.2%; Pred. No. 0.035;
Matches 130; Conservative 48; Mismatches 210; Indels 172; Gaps 26;

QY 19 GQQGFAIPIGQAMAIAGQIRSGGSPVHIG--PTAFGLGWEDN--NGGARVQRVVGS 74
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 244 GQQQPGVPLGYPIK-APKPGYGLPTN-CKLPYGVAGAGKAGYPTGTGVSQAATAA 301
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 75 APAAS-LGISTGDVITAVDGAIPINSATAMADALNGHHHPGDVISVTWTKSGGTRTGNVT 133
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 302 AKAAYAGAGGGVLPVGGGGIFGGAGAIPIGG-----ITGACTPAAAAAAK 350
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 134 AEGPPAEF-----LVPRGMSRA-----FIIDPTISAIIDGLYDLGI-GIPNOGGILYS 181
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 351 AAKAAKYGAAGGLVPGGPGVRVPCAGIPGVIGPVGGIPGVGGIPGVGG----- 406
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 182 SLEVF EKALEBELAAAFPGDGMWG-----SAADKYAGKNNHNVFFQELADLDRQL 231
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 407 -----PGIGGPGIVGGPCAVSPAAAAKAAKAAKY----- 436
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 232 ISLIHQANAVQTTREDILEGAKKGLFVVRPVAVDLTYIPVVGHALSAAFPQPCAGAMAV 291
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 437 -----GARGGVG-----IPTYG--VGAGGPPGYGVGAGAG 464
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 292 VGGALAVLVKTLINATOLLKLAELVAAAIADIISDVADIKIGILGEVWEFITNAL 351
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 465 LGGA-----SQAAAAAA-----AKAAKYGAGGAGTGLGLVPGAV 499
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 352 NGLKELMDKLTGWVTGLFSRGSNLESFFAGVPG-LTGATSGLSQVTLGFGAAGLSASSG 410
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 500 PG-----ALPGAVFGALP-----GAVPGALPGAVFPGTGGVPCAGTPAAAAA 543
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 411 LAHADSLASSASL---PALAGIGGSGFGGLP---SLAQVHAASL--RQALRRADGPVG 462
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 544 AAKAAKAAKAGQYGLPGVGVPVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGG 600
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 463 AAAEQVGQSQSLVSAQSSQ-----GNVGGP-VGVGMHPSSGA-----SKGTTTKKYSE 509
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 601 GAGTPAAAKSAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 660
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

QY 510 GAAAGTDAERAPVEADAGG 529
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 661 GAVPGSLAASKAAKYGAAGG 680
   |||  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

```

Search completed: November 21, 2003, 16:09:59
Job time : 13.2444 secs

This Page Blank (uspto)

GenCore version 5.1.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 6.35556 Seconds
 (without alignments)
 3988.226 Million cell updates/sec

Title: US-09-688-672A-64

Perfect score: 2737

Sequence: 1 MHHHHTAASDNFOLSQGG.....RAPVEADAGGQKVLVRNVV 539

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwisProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	173	6.3	836	1 VG26 BPMD2	O64220 mycobacteri
2	165	6.0	864	1 ELS_RAT	Q93372 rattus norv
3	164	6.0	836	1 VG26 BPMD5	Q05233 mycobacteri
4	159.5	5.8	860	1 ELS_MOUSE	P54320 mus musculu
5	152.5	5.6	730	1 ELS_HUMAN	P15502 homo sapien
6	152.5	5.6	914	1 WAZ2 MYCTU	O05794 mycobacteri
7	150	5.5	801	1 Y747 MYCTU	Q53810 mycobacteri
8	147	5.4	960	1 DUG1 DROME	P31007 drosophila
9	146.5	5.4	515	1 Y140 MYCTU	Q50594 mycobacteri
10	146	5.3	957	1 Y278 MYCTU	P56877 mycobacteri
11	143.5	5.2	2333	1 PGCA CANFA	Q28343 canis fami
12	142.5	5.2	2124	1 PGCA RAT	P07897 rattus norv
13	142	5.2	778	1 Y034 MYCTU	P19333 mycobacteri
14	142	5.2	1901	1 Y208 MYCTU	Q53553 mycobacteri
15	141	5.2	2132	1 PGCA MOUSE	Q61282 mus musculu
16	139	5.1	435	1 YU21 MYCTU	Q53268 mycobacteri
17	139	5.1	543	1 YP91 MYCTU	Q50630 mycobacteri
18	138.5	5.1	419	1 YC58 MYCTU	Q11060 mycobacteri
19	137	5.0	2038	1 FSH DROME	P13709 drosophila
20	136	5.0	747	1 ELS_BOVIN	P04985 bos taurus
21	136	5.0	2415	1 PGCA HUMAN	P16112 homo sapien
22	135.5	5.0	463	1 Y468 MYCTU	Q53416 mycobacteri
23	135.5	5.0	557	1 PAC1 PSESV	Q05053 pseudomonas
24	135.5	5.0	750	1 ELS_CHICK	P07916 gallus gall
25	133.5	4.9	408	1 HTRA LACLA	Q91A06 lactococcus
26	131.5	4.8	558	1 YJ83 MYCTU	Q10873 mycobacteri
27	130	4.7	498	1 Y118 MYCTU	Q50615 mycobacteri
28	130	4.7	3591	1 FRAB_BORPE	P12255 bordetella
29	128.5	4.7	603	1 YD25 MYCTU	Q10637 mycobacteri
30	127.5	4.7	1156	1 GLH4 CAEBL	O76743 caenorhabdi
31	125.5	4.6	557	1 PAC1 PSESV	P15557 pseudomonas
32	125.5	4.6	1758	1 CA24 CAEBL	P17140 caenorhabdi
33	125	4.6	413	1 HTRA_LACHE	Q924h7 lactobacill

34 124.5 4.5 804 1 BCB3 ACEXY Q9WX62 acetobacter
 35 124.5 4.5 1763 1 CA24_ASCSU P27393 ascaris suu
 36 124.5 4.5 2243 1 OMPA_RICRI P15921 rickettsia
 37 124 4.5 5263 1 PROH_BOMBO P05790 bombyx mori
 38 123.5 4.5 747 1 SPDI_NBPCL P19837 nephila cla
 39 123 4.5 306 1 EC40 DAUCA Q07322 daucus caro
 40 122.5 4.5 1372 1 CA21 MOUSE Q01149 mus musculu
 41 121.5 4.4 842 1 PLEC CAUCR P37894 caulobacter
 42 121.5 4.4 1298 1 ICP4_HSV1 P08392 herpes simp
 43 121.5 4.4 1691 1 CA64_HUMAN Q14031 homo sapien
 44 121 4.4 1356 1 CA24_ONCMY O93484 oncornynchu
 45 120.5 4.4 689 1 YO27_BPHPI P51731 bacteriophu

ALIGNMENTS

RESULT 1
 VG26_BPMD2
 ID VG26_BPMD2 STANDARD; PRT; 836 AA.
 AC O64220;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Minor tail protein GP26.
 GN 26.
 OS Mycobacteriophage D29.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=28369;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98300335; PubMed=9636706;
 RA Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
 RT "Genome structure of mycobacteriophage D29: implications for phage evolution.";
 RL J. Mol. Biol. 279:143-164 (1998).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF022214; AAC18467.1; -
 DR PIR; H72802; H72802.
 FT INIT MET 0 0 BY SIMILARITY.
 SQ SEQUENCE 836 AA; 86572 MW; 528D12ED0FCB92E6 CRC64;

Query Match 6.3%; Score 173; DB 1; Length 836;
 Best Local Similarity 19.8%; Pred. No. 0.011;
 Matches 136; Conservative 94; Mismatches 245; Indels 212; Gaps 25;
 QY 17 SQGGQGFAPIG-----QAMAIAGQIRSGGSPTVHIG----- 50
 Db 65 TEARAGVNIPTDVNDKDTKGFLSRLGKGKLSLGDAAKASQVQHLGKSFLLDTR 124
 QY 51 TAFELGVVDNNGNGARVORVG-----SAPA--ASLGISTGVDVITAVDGAPINSATA 101
 Db 125 TAWIGVGIV-----AIAAPLVGLVAGLGLPILLSAFGAGVGVVALGMDGIKA-AAET 177
 QY 102 MADALNGHPGDVISTWTOT-----KSGG-----TRTGNVTIAEGPPAEFLV 143
 Db 178 MPALAE--AKTAVSTFTQGLTPVQQQLGGLLTITPNLQNVATGIVNIAGFTDVVSQ 235
 QY 144 PRGMSRAFIPTTISAIDGLYDLIGIG-----PNQG-----GLYSSLYFEKALEEL 193
 Db 236 GPGLQQLQNLDRGTGFFTGGLGVISTGTAFLITLNAGANAFGLLAPLQSFANGFNDM 295
 QY 194 AAAPFGDGLGSAADYACKNKNHNFQELAD----- 226
 Db 296 VNRVTNGVDFDGANQGLSQTILGNLNFRLMESGLQAMQGLGGPLSTLVNGIGDLFIAL 355

DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PR01500; TROPELASTIN.
KW Structural protein; Repeat; Signal; Connective tissue.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 860 ELASTIN.
FT DISULFID 850 855 BY SIMILARITY.
SQ SEQUENCE 860 AA; 71955 MW; 0C0BESAEEEDD7F1 CRC64;

Query Match 5.8%; Score 159.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 0.063;
Matches 118; Conservative 38; Mismatches 186; Indels 165; Gaps 22;

QY 19 GGGGFAIPGQAWIAAGQIRSGGSGTTHVG--PTAFGLGLGVVDN--NNGNARVQRVVG 74
DQ 19 GGGGFAIPGQAWIAAGQIRSGGSGTTHVG--PTAFGLGLGVVDN--NNGNARVQRVVG 74
DB 259 GGGGPGVPLGPIK-APKLPGGVGLPTN-GKLPYGVAGAGKAGYPTGTGVGSGQAAAA 316
QY 75 APAASLGISTGVDITAVDGAIPNSATAMADALNGHPGDVIVTWQTKSGGTRTGNVTIA 134
DB 75 APAASLGISTGVDITAVDGAIPNSATAMADALNGHPGDVIVTWQTKSGGTRTGNVTIA 134
DB 317 AKAAKYGAGAGVLPVGGGIPGGAGATPGIGG-----IAGATPAAAAAKAAKAA 370
QY 135 EGPPAFBLVPRGMSR---AFITDPTISAIDGLYDLGICIPNOGGILYSSLEYFEKALE 191
DB 135 EGPPAFBLVPRGMSR---AFITDPTISAIDGLYDLGICIPNOGGILYSSLEYFEKALE 191
DB 371 KYGAAGLVPGGPGVRLPGAGI--PGVGGIPGVGGIPGVGGPGIGG-----414
QY 192 ELAAATPGDGLG--SAADKYAGKRNHNVFFQELADLDRLQLISLHDAQNAVQTTTRDILE 250
DB 192 ELAAATPGDGLG--SAADKYAGKRNHNVFFQELADLDRLQLISLHDAQNAVQTTTRDILE 250
DB 415 --PGIVGGPGAVGAPAAAKAAKAAKAAK---439
QY 251 GAKKGLFVFPVAVDLTYIPVGHASAAFAQPFACAGAMVVGALAYLVKTLINATOL 310
DB 251 GAKKGLFVFPVAVDLTYIPVGHASAAFAQPFACAGAMVVGALAYLVKTLINATOL 310
DB 440 GARGVG-----IPTYG--VGAGGFPVGVGAGAGLGA-----SPAA 475
QY 311 LKLLAKLAELVAAAIADIISVDADIKGLGEVWEFTITNALNGLKELWKLWGTGTFGS 370
DB 311 LKLLAKLAELVAAAIADIISVDADIKGLGEVWEFTITNALNGLKELWKLWGTGTFGS 370
DB 476 AAAAAKAAKYGAGG-----AGALGGL-----VFGAVPG---503
QY 371 RGWSNLEFPAGVPGLTGATSGLSQVTLGFLGAAGLSAGSLAHADSLAGSL--PALAGI 429
DB 371 RGWSNLEFPAGVPGLTGATSGLSQVTLGFLGAAGLSAGSLAHADSLAGSL--PALAGI 429
DB 504 -----ALPGAVPAVPGAGVPG-----AGTFAAAAAAKAAKAAKAGLFGVGVG 548
QY 430 GGGSGFGGLFSLAQVHAASQALRPRADPGVPGAAAEQVGGSQLVSAQSGQMG-GPVG 488
DB 430 GGGSGFGGLFSLAQVHAASQALRPRADPGVPGAAAEQVGGSQLVSAQSGQMG-GPVG 488
DB 549 PGGVGGVGP-----GGVG-----VGGVPGGVGGVGTGIGAGPGG 584
QY 489 MGM-HPSSGASKGTTTKYSEGAAG 514
DB 489 MGM-HPSSGASKGTTTKYSEGAAG 514
DB 585 LGGAGSPAAAKSAKAAKAAKAAKAAKAAAG 611

RESULT 5
ELS_HUMAN
ID ELS_HUMAN STANDARD; PRT; 730 AA.
AC P15502; Q14233; Q14238;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Elastin precursor (Tropoelastin).
GN ELN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN R1
RP MEDLINE=87289668; PubMed=3039501;
RA Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA Rosenbloom J.C., Peltonen L., Rosenbloom J.,
RT "Alternative splicing of human elastin mRNA indicated by sequence
analysis of cloned genomic and complementary DNA."
RL Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684 (1987).
[2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Skin fibroblast.

RA Fazio M.J., Olsen D.R., Kauh E.A., Baldwin C.T., Indik Z.,
RA Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.,
RT "Cloning of full-length elastin cDNAs from a human skin fibroblast
recombinant cDNA library: further elucidation of alternative splicing
utilizing exon-specific oligonucleotides."
J. Invest. Dermatol. 91:458-464 (1988).
[3]
RN R1
RP SEQUENCE OF 164-724 FROM N.A. (ISOFORM B).
RC TISSUE=Placenta;
MEDLINE=98156138; PubMed=2831431;
RA Fazio M.J., Olsen D.R., Kuivaniemi H., Chu M.L., Davidson J.M.,
RA Rosenbloom J., Uitto J.,
RT "Isolation and characterization of human elastin cDNAs, and age-
associated variation in elastin gene expression in cultured skin
fibroblasts."
Lab. Invest. 58:270-277 (1988).
[4]
RN R1
RP SEQUENCE OF 603-730 FROM N.A.
RC TISSUE=Hippocampus, and Placenta;
MEDLINE=96291399; PubMed=8689688;
RA Frangiskakis J.M., Ewart A.K., Morris C.A., Mervis C.B.,
RA Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
RA Green E.D., Roessel C., Gutowski N.J., Noble M., Atkinson D.L.,
RA Odelberg S.J., Keating M.T.,
RT "LIM-kinase hemizygosity implicated in impaired visuospatial
constructive cognition."
Cell 86:59-69 (1996).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
INTO AN EXTENSIBLE 3D NETWORK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Comment=Additional isoforms seem to exist;
Name=1;
IsoId=P15502-1; Sequence=Displayed;
Name=2;
IsoId=P15502-2; Sequence=VSP_004243;
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -!- DISEASE: Haploinsufficiency of ELN may be the cause of certain
cardiovascular and musculo-skeletal abnormalities observed in
Williams-Beuren syndrome (WBS), a rare developmental disorder. It
is a contiguous gene deletion syndrome involving genes from
chromosome band 7q11.23.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M17282; AAC98394.1; JOINED.
EMBL; M16983; AAC98394.1; JOINED.
EMBL; M17265; AAC98394.1; JOINED.
EMBL; M17266; AAC98394.1; JOINED.
EMBL; M17267; AAC98394.1; JOINED.
EMBL; M17268; AAC98394.1; JOINED.
EMBL; M17270; AAC98394.1; JOINED.
EMBL; M17271; AAC98394.1; JOINED.
EMBL; M17272; AAC98394.1; JOINED.
EMBL; M17273; AAC98394.1; JOINED.
EMBL; M17275; AAC98394.1; JOINED.
EMBL; M17276; AAC98394.1; JOINED.
EMBL; M17277; AAC98394.1; JOINED.
EMBL; M17278; AAC98394.1; JOINED.
EMBL; M17279; AAC98394.1; JOINED.
EMBL; M17280; AAC98394.1; JOINED.
EMBL; M17281; AAC98394.1; JOINED.
EMBL; M36860; AAA52382.1; -


```
QY 134 AEGPPAEFLVPRGMSRAFIIDPTISAIIDGLYDLGLGIPNOGGILYSSLEYFEKALEBL 193
Db 462 AGG-----NAV-----LIGN-GEGLI-----GAL 480
QY 194 A--AAPGDLGWSAADKYAGRNHNVPFQELADLDROLISLIHQANAVOTTRDIL-- 249
Db 481 AKSGFGGFGGLLGDGY-----NAPESTSPMHNLOQDILFNEPTAL-TGRLIGN 534
QY 250 -----EGAKKGLFVRPVAVDLTVPVVGHLSAFAFPFCAGAMAVVUGALAYL 299
Db 535 GDSGTCTGDDGAGGWLFP-----GNGGNGGAGNAGNGSAGGAGGGL 580
QY 300 VVKTLINATOLLKLAELVAAAIADIIISVDADIIKIGLVEWFTFNALNGKELMD 359
Db 581 -----FTGGAGGAGGVTAGAGGAGGAGGSAFLIGSGGTGGVGA-A 622
QY 360 KLTGVTGLFSRGSNLESFFAGVPG-----LTGATSGLSQVTLFGAAGLSAGSLAH 413
Db 623 TTTGGVGG--AGNAGLIGAGLGGCGGAFAGVTGTGAGGTG--GAAGLFANGGAGG 678
QY 414 ADSLASSASLPALAGIGGSGFGGLPFLSLAQVHAASRQALRPADGFGVGAEEQ----- 467
Db 679 AGGTGTA-----GGAGGAGGAGL-----VAHGCT-----GGPGNGSGTGAGGTG 720
QY 468 -VGQSLVSAQSGSQMGPGVGMGHPSPGASKGTTTKYSEGAAGTDEAPVAD 526
Db 721 GAGGPGGLYAGGSGGAGGGMAGGGVGGNAGSLTLNASGG--AGSGGSSLSGKAG 778
QY 527 AGG 529
Db 779 AGG 781

RESULT 7
Y747 MYCTU
ID Y747 MYCTU STANDARD; PRT; 801 AA.
AC OS3810;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Hypothetical PE-PGRS family protein Rv0747 precursor.
GN Rv0747 OR MT0772.5 OR MT041.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterlinden T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
```

```
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGSS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL021958; CA117514.1; -
CC EMBL; AE006988; AAK45011.1; -
CC PIR; F70824; F70824.
CC TIGR; MT072.5; -
CC Tuberculast; Rv0747; -.
CC InterPro; IPR000084; PE_1_region.
CC Pfam; PF00934; PE; 1_region.
CC Hypothetical protein; Repeat; Signal; Complete proteome.
KW SIGNAL 1 30
FT CHAIN 31 801
FT CONFLICT 188 188 G -> S (IN REF. 2).
FT CONFLICT 225 225 R -> G (IN REF. 2).
FT CONFLICT 227 227 R -> G (IN REF. 2).
FT CONFLICT 295 295 K -> R (IN REF. 2).
FT CONFLICT 300 300 S -> G (IN REF. 2).
FT CONFLICT 338 338 T -> I (IN REF. 2).
FT CONFLICT 377 377 A -> P (IN REF. 2).
FT CONFLICT 577 577 T -> A (IN REF. 2).
FT CONFLICT 580 595 MISSING (IN REF. 2).
SQ SEQUENCE 801 AA; 65407 MW; EA54C9BF45A00F41 CRC64;

Query Match 5.5%; Score 150; DB 1; Length 801;
Best Local Similarity 28.1%; Pred.No.0.2;
Matches 72; Conservative 23; Mismatches 111; Indels 50; Gaps 10;

QY 275 ALSAFAQAPFCAGAMAVVGGALAYLVVKTLLNATQLKLAKL-AELVAAAIADIISDVA 333
Db 67 AQAAAFYAQF-VQALSAGGAYA-----AAEAAVSPLLAPINAQFVATGRPLINGA 119
QY 334 DIKIGLGEVNEFTNALNGKELMDKLTQWVTGLFSRGSNLESFFAGVPGTGTATSG 393
Db 120 NGAPG-----TGANGG-----PGWMLICNGGAGGS-----GAPGAGAGNG- 155
QY 394 SQVTGLFMAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPFLSLAQVHAASRQAL 453
Db 156 -----GAGGLFGSGG-----AGASTDVAGGAGGAGGAGG--NAGMLFGAAGVGGV 199
QY 454 RPRADGPGVAAAEQVQSQSLVSAQSGQMGPGVGMGHPSPGASKGTTTKYSEGA 513
Db 200 GGFSGNGATGAGGAGGAGGLFGAGRERGSGSGNLTGGAGGAGNAGTTLAT--GDGAG 257
QY 514 GTEDAERAPVEADAGG 529
Db 258 GTGASRSRSGFGGAGG 273

RESULT 8
DLGL DROME
ID DLGL DROME STANDARD; PRT; 960 AA.
AC P31007;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Discs large-1 tumor suppressor protein.
GN DLGL OR L1/DLGL
OS Drosophila melanogaster (Fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
```


Best Local Similarity 28.4%; Pred. No. 0.41;		Matches 89; Conservative 17; Mismatches 119; Indels 88; Gaps 15;	
QY	257 EFVRPVAVDITYPVCHALSAFAQP- - - - -FCAG- - - - -ANAVVGA- - - - -LAVLV 301		
Db	9 EVTAAATDIASL- - - - -GSSISAANAANAANTTALMAAGADEVSTAALFAGHQAYQAL 65		
QY	302 KTLINA- - - - -TOLLKLLAKLAEVAAIAADIIISDVADIKIGILGEVWF- - - - -ITNA 350		
Db	66 SAQAQAFHAFVQALTSGGAYAAAEAAVPLDPN- - - - -EFFLANTGRPLING 118		
QY	351 LNLKELNDKLT- - - - -CWVTGLSRGWSNLESFAGVPGTGTATSLGSLQVTLFGAA 403		
Db	119 ANAPG- - - - -TGANGDGMWLTNGGAGGS- - - - -GAAGYNGAGNGGAGNGGAG 166		
QY	404 GLSASSGLAHADSLASSASLPALAG- - - - -GCGSGFGGLPSLAQVHAASR- - - - -QALRPADG 459		
Db	167 GLTNGGAGGAGGVASS- - - - -GIGSGAGGNAMLFAGGAGGAGGVVALTGGAG 219		
QY	460 PVGA- - - - -AAEQVGGQ- - - - -SOLVSAQSGQMGFVGMGMHPSGASKGTTTK 505		
Db	220 AGGAGGNAGLLFGAAGYGGAGGTNGSALGGAGGAGGLFATGTVGGSGGAG- - - - - 273		
QY	506 KYSEGAAGTEDA 518		
Db	274 - - - - -SSGGAGGAGGA 284		
RESULT 11			
PGCA CANFA			
ID	PGCA CANFA	STANDARD;	PRT; 2333 AA.
AC	Q28343; Q28310;		
DT	0-NOV-1997 (Rel. 35, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).		
GN	AGC1.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RA	Glant T.T., Adams M.E., Kwok S.X.F., Huang D., Fuellep C.;		
RT	"Complete coding sequence and deduced amino acid sequence of aggrecan of canine cartilage."		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE OF 774-833 FROM N.A.		
RC	TISSUE=Cartilage;		
RX	MEDLINE=9512852; PubMed=7827755;		
RA	Barry F.P., Neame P.J., Sasse J., Pearson D.;		
RT	"Length variation in the keratan sulfate domain of mammalian aggrecan."		
RL	Matrix Biol. 14:323-328(1994).		
RN	[3]		
RP	SEQUENCE OF 1830-2333 FROM N.A.		
RA	Adams M.E., Kwok S.X.F., Huang D., Glant T.T., Fullop C.;		
RL	Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE OF 2082-2118 FROM N.A.		
RC	TISSUE=Cartilage;		
RX	MEDLINE=9335252; PubMed=8349621;		
RA	Fuellep C., Walcz E., Valyon M., Glant T.T.;		
RT	"Expression of alternatively spliced epidermal growth factor-like domains in aggrecans of different species. Evidence for a novel module."		
RL	J. Biol. Chem. 268:17377-17383(1993).		
CC	-!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A		
CC	REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.		
CC	-!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).		
CC	-!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.		
CC	-!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).		
CC	-!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.		
CC	-!- SIMILARITY: Contains 4 link domains.		
CC	-!- SIMILARITY: Contains 1 C-type lectin family domain.		
CC	-!- SIMILARITY: Contains 1 Sushi (SCR) domain.		
CC	-!- SIMILARITY: Contains 1 EGF-like domain.		
CC	-!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	EMBL; U65989; AAB06238.2; -		
DR	EMBL; S74662; AAC60527.1; -		
DR	EMBL; L07054; -; NOT_ANNOTATED_CDS.		
DR	PIR; I46998; I46998.		
DR	KSSP; P08709; IBP9.		
DR	InterPro; IPR000152; Asx hydroxyl.		
DR	InterPro; IPR000742; EGF 2.		
DR	InterPro; IPR001881; EGF Ca.		
DR	InterPro; IPR006209; EGF-like.		
DR	InterPro; IPR007110; Ig-like.		
DR	InterPro; IPR003006; Ig_MHC.		
DR	InterPro; IPR001304; Lectin_C.		
DR	InterPro; IPR000538; Link.		
DR	InterPro; IPR003324; SGXSG.		
DR	InterPro; IPR000436; Sushi_SCR_CCP.		
DR	Pfam; PF00008; EGF; 1.		
DR	Pfam; PF00047; Ig; 1.		
DR	Pfam; PF00059; Lectin_C; 1.		
DR	Pfam; PF02339; SGXSG; 66.		
DR	Pfam; PF00084; sushi; 1.		
DR	Pfam; PF00193; Xlink; 4.		
DR	PRINTS; PR01265; LINKXODULE.		
DR	PRINTS; PR00356; ANTIFREEZII.		
DR	ProDom; PD000918; Link; 4.		
DR	SMART; SM00032; CCP; 1.		
DR	SMART; SM00034; CLECT; 1.		
DR	SMART; SM00179; EGF_CA; 1.		
DR	SMART; SM00445; LINK; 4.		
DR	PROSITE; PS00010; ASX HYDROXYL; 1.		
DR	PROSITE; PS00615; C TYPE LECTIN 1; 1.		
DR	PROSITE; PS00041; C TYPE LECTIN 2; 1.		
DR	PROSITE; PS00022; EGF_1; 1.		
DR	PROSITE; PS01187; EGF_CA; 1.		
DR	PROSITE; PS00835; IG LIKE; 1.		
DR	PROSITE; PS00290; IG_MHC; 1.		
DR	PROSITE; PS01241; LINK; 4.		
KW	Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;		
KW	EGF-like domain; Repeat; Immunoglobulin domain.		
FT	POTENTIAL.		
FT	SIGNAL 1 16		
FT	CHAIN 17 2333		
FT	DOMAIN 34 147		
FT	DOMAIN 170 247		
FT	DOMAIN 268 349		
FT	DOMAIN 513 590		
FT	DOMAIN 611 652		
FT	DOMAIN 2081 2117		
FT	EGF-LIKE, CALCIUM-BINDING (POTENTIAL).		

FT DOMAIN 2130 2245 C-TYPE LECTIN.
 FT DOMAIN 2249 2307 SUSHI.
 FT DOMAIN 248 140 GI-A.
 FT DOMAIN 152 247 GI-B.
 FT DOMAIN 253 349 GI-B'.
 FT DOMAIN 495 589 G2-B.
 FT DOMAIN 596 691 G2-B'.
 FT DOMAIN 694 816 KS.
 FT DOMAIN 819 1334 CS-1.
 FT DOMAIN 1395 2079 CS-2.
 FT DOMAIN 2080 2333 G3.
 FT DISULFID 51 133 BY SIMILARITY.
 FT DISULFID 175 246 BY SIMILARITY.
 FT DISULFID 199 220 BY SIMILARITY.
 FT DISULFID 273 348 BY SIMILARITY.
 FT DISULFID 297 318 BY SIMILARITY.
 FT DISULFID 518 589 BY SIMILARITY.
 FT DISULFID 542 563 BY SIMILARITY.
 FT DISULFID 616 691 BY SIMILARITY.
 FT DISULFID 640 661 BY SIMILARITY.
 FT DISULFID 2085 2096 BY SIMILARITY.
 FT DISULFID 2090 2105 BY SIMILARITY.
 FT DISULFID 2107 2116 BY SIMILARITY.
 FT DISULFID 2123 2134 BY SIMILARITY.
 FT DISULFID 2151 2243 BY SIMILARITY.
 FT DISULFID 2219 2235 BY SIMILARITY.
 FT DISULFID 2250 2293 BY SIMILARITY.
 FT DISULFID 2279 2306 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 676 676 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 2333 AA; 240573 MW; 889ED78F3508B596 CRC64;

Query Match 5.2%; Score 143.5; DB 1; Length 2333;
 Best Local Similarity 21.6%; Pred. No. 1.5;
 Matches 127; Conservative 60; Mismatches 183; Indels 219; Gaps 27;

QY 41 GGSPT-----VHIGTAPLGLGVNDNGNGARVORVVGSAASLGISTGDIVITAV 91
 DB 706 GGTTPSVVEDWPTQVGP-----VPSVP-----MGEITAI 738
 QY 92 DGAPI--NSATAMADALN--GHEPGDIVSVTQKSGGTRTGNVTLAERG----- 137
 DB 739 LDFTIEPENQTEWEPAYSPAGTSPLPGIPPTWPTTSTATEES---TEGSPGTEVPVSVE 794
 QY 138 ---PAEFLVPGMSRPAIIDTISADGLYDLGLI-----GIPNOGGTLY 180
 DB 795 BPSPEEPPEWELSTLSPGSPGTELPGSGEASGVPEVSGDFTGSGEVSGHPDSSGOLS 854
 QY 181 SSLEYFEKALELAAAFPGDGLWGAADKYAKGNHNVFFQELADLDRQLISL----- 234
 DB 855 G-----BSASGLPSEDLDSGLTSVAVSG-----LASGDEDRITLSSIPKV 895
 QY 235 ----IHDQNAVQTRDILEGAKKLEFVRPVAVDLTYIPVGHALSAAFOAPCAGAMA 290
 DB 896 EGEGLTASGVEDLSGLPSPG-REGLETSTSGVGLSLP-SGEGL----- 939
 QY 291 VVGALAVLWKTLLINATQLLKALAEVLAADIAI-----TSDVADIKG 338
 DB 940 -----EVSASGVEDLSGLPSPGPGPTSTSGVGLSRL 971
 QY 339 ILGEVWEFTNALNGKELWKLTGVTGLPSRGNLSFAGVGLTG-----A 389
 DB 972 PSGEPEVSASGV-----GELSGLPS-GREGLETSTSGVEDLSGLPSPGGEPEAS 1019
 QY 390 TSGLSQVTGL-----FGAAGLSASSGLAHADSLASSAS-LPALAGIGGG-----SG 434
 DB 1020 TSGVGLSLRPLPSGEPEVSASGVEDLSGLPSPGGELEASGVGLSLPSPGGEPEASAG 1079

QY 435 FGGLPSLAQVHRASTRQALRPADGP-VGAAAEQ-----VGGSQLVSAGSQMGPG-PV 487
 DB 1080 VGLSLRL-----PSGEGPEVSASGVEDLSGLSSGESPEASASGVGLSLGLPS 1126
 QY 488 GMGMHPSS---GASKGTTTKYSEGAAAGTDAERAP---VEADAGG 529
 DB 1127 GREGLTASGVGLSLGLPSGEGQEASGVEDLSRLPSGEGPEASAG 1175

RESULT 12
 PGCA RAT
 ID PGCA RAT STANDARD; PRT; 2124 AA.
 AC P07897;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 GN AGC1 OR AGC.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88087070; PubMed=3693370;
 RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
 RT "Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones";
 RL J. Biol. Chem. 263:17757-17767(1987).
 RN [2]
 RP REVISION TO 698.
 RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
 RL J. Biol. Chem. 263:10040-10040(1988).
 RN [3]
 RP SEQUENCE OF 1856-2124 FROM N.A.
 RX MEDLINE=8250698; PubMed=2424893;
 RA Doerge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;
 RT "Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan";
 RL J. Biol. Chem. 261:8108-8111(1986).
 CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
 CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
 CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 4 link domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M13518; AAA41836.1; -;
DR EMBL; J03485; AAA21000.1; ALT_SEQ.
DR PIR; A92623; A28452.
DR HSSP; P98066; 1TSG.
DR InterPro; IPR002353; AntifreezeII.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig MHC.
DR InterPro; IPR003596; Ig_v.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR003324; SGXSG.
DR InterPro; IPR000436; Sushi_scr_CCP.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF02333; SGXSG; 55.
DR Pfam; PF02333; SGXSG; 55.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 4.
DR PRINTS; PR01265; LINKMODULE.
DR PRINTS; PR00356; ANTIFREEZEII.
DR ProDom; PD000918; Link; 4.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00406; IGV; 1.
DR SMART; SM00445; LINK; 4.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS01241; Link; 4.
DR Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW Repeat; Immunoglobulin domain.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2124 AGGREGAN CORE PROTEIN.
FT DOMAIN 34 147 IG-LIKE V-TYPE.
FT DOMAIN 170 247 LINK 1.
FT DOMAIN 268 349 LINK 2.
FT DOMAIN 504 581 LINK 3.
FT DOMAIN 602 683 LINK 4.
FT DOMAIN 1910 2036 C-TYPE LECTIN.
FT DOMAIN 2040 2098 SUSHI.
FT DOMAIN 48 140 Gl-A.
FT DOMAIN 152 247 Gl-B.
FT DOMAIN 253 349 Gl-B'.
FT DOMAIN 486 580 G2-B.
FT DOMAIN 587 682 G2-B'.
FT DOMAIN 685 798 KS.
FT DOMAIN 801 1226 CS-1.
FT DOMAIN 1227 1909 CS-2.
FT DOMAIN 1910 2124 G3.
FT DISULFID 51 133 BY SIMILARITY.
FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 509 580 BY SIMILARITY.
FT DISULFID 533 554 BY SIMILARITY.
FT DISULFID 607 682 BY SIMILARITY.
FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 1914 1925 BY SIMILARITY.
FT DISULFID 1942 2014 BY SIMILARITY.
FT DISULFID 2010 2026 BY SIMILARITY.
FT DISULFID 2041 2084 BY SIMILARITY.
FT DISULFID 2070 2097 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1842 1842 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2124 AA; E30BBE61593A34B1 CRC64;
5.2%; Score 142.5; DB 1; Length 2124;

Query Match

Best Local Similarity 21.9%; Pred. No. 1.6;
Matches 126; Conservative 75; Mismatches 194; Indels 181; Gaps 31;
QY 24 AIPGQMAIAGCIRSGGSPVTHICPTAFGLGVVDNNGARVQRVVGSAAPASLGIS 83
Db 796 AVPSGTELPSSGDT--SGAPDLSGDFT----GSTDTSG--RLDSSGEPGSGESGLE 844
QY 84 TGDV-----ITAVDGAIPNSATMADALNGHHGPDVSVTWQTSKGGTRTGNVLAEGP 137
Db 845 SGDLSSGLGPTVSGLPVSGSASGDG-----EIPWSS-----T 879
QY 138 PAEFLVPRGSMGRAFIIDPTISAIIDGLYDILLGIPNOGILYSSLEYFEKALEELAAAF 197
Db 880 PTVDRLPSGGS-----LEGSASA-SGTGL--SGLPSCGEITETS-----ASGTEISG-- 926
QY 198 PGDGLWSAADKYAGKNNHNVNFOELADLDQLISLIHQANAVOTTRDILEGAKKGLE 257
Db 927 ----LPSGGD-----DLETSTSGI--DGASVLPTRGGLETSSASGVE 962
QY 258 FVRPV-----AVDLTYIPVVGHALSAFAAFPCAGAMAVVGALAYLVVKTL 304
Db 963 DLGSLPSGEGSETSTSGIEDISVLTGESPTSGVGDLSGLPS--GGESLETSSASGV 1020
QY 305 INATQLKLAKLAELVAAAIAI-----ISDVADIKIGILGEVWEFT 348
Db 1021 EDVTQLPTEGGI-ETSASGIEDITVLTGRENLETSSASGVEDVSLPSGKEG----LE 1074
QY 349 NALNGKLKELWDKLTGVTGLFSRGHNSLES-----FFAGVP-----CLTGATSLGSOVTGL 399
Db 1075 TSASGIEDI-----SVFPTAEGLSETSSAGGVSGIPSGEDGTETSTSGVEGVSGL 1125
QY 400 -FGAAGLSASS-----GLAHADSLASSALPALAGI-----GGSGSGFGGLPS 440
Db 1126 PSGEGLETSSASGVEDLGLPTSDLETSSASGVDTGTPSGREDTETSPGVGDLSGLP- 1184
QY 441 LAQVHAASRQALRRADPGVGAABQVG-----GGSQLV-SAGCSGGMG-----GPVGMGG 491
Db 1185 -----SQEGLETSSASG-----AEDLGLPSGKEDLVGSAGALDFGKLPSTGLSGQ 1232
QY 492 MHPSSGASKGTTTKYS-----EGAAAGTEDAERAP 522
Db 1233 TPEASGLPSG-FSGEYSGVDIGSGPSSGLPFDPSGLP 1267
RESULT 13
YQ34 MYCTU
ID_YQ34_MYCTU STANDARD; PRT; 778 AA.
AC P71933;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein Rv2634c.
GN Rv2634C OR MT2712 OR MTCV441.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL Nature 393:537-544(1998).
RN [2]

RESULT 15

PGCA_MOUSE
 ID PGCA_MOUSE STANDARD; PRT; 2132 AA.
 AC Q61282; Q64021;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 DE AGC1 OR AGC.
 GN AGC1 OR AGC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Cartilage;
 RX MEDLINE=9510487; PubMed=7806222;
 RA Walcz E., Deak F., Erhardt P., Coulter S.N., Fuellep C., Horvath P., Doege K.J., Glant T.J.;
 RA "Complete coding sequence, deduced primary structure, chromosomal localization, and structural analysis of murine aggrecan.";
 RL Genomics 22:364-371(1994).
 [2]
 RP SEQUENCE OF 211-326 FROM N.A.
 RC STRAIN=129/Sv;
 RX MEDLINE=95004579; PubMed=7920633;
 RA Watanabe H., Kinata K., Line S., Strong D., Gao L.-Y., Kozak C.A., Yamada Y.;
 RA "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene.";
 RL Nat. Genet. 7:154-157(1994).
 [3]
 RP INTERACTION WITH FBLN1.
 RX PubMed=16400671;
 RA Asberg A., Adam S., Kostka G., Timpl R., Heinegaard D.;
 RA "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and versican.";
 RL J. Biol. Chem. 274:20444-20449(1999).
 CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -!- SUBUNIT: INTERACTS WITH FBLN1.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
 CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
 CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -!- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF CARTILAGE MATRIX DEFICIENCY (CMD), AN AUTOSOMAL RECESSIVE SYNDROME CHARACTERIZED BY CLEFT PALATE, SHORT LIMBS, TAIL AND SNOOT. MUTATION IN STRAIN CMD CAUSES ABSENCE OF AGGREGAN BY TRUNCATION OF THE PROTEIN (MUTATION IN THE G1 DOMAIN).
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 4 link domains.
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; L07049; AAC37670.1; .
 CC DR EMBL; S73722; AAB32160.1; .
 CC DR EMBL; S73721; AAB32160.1; JOINED.
 CC DR PIR; A55182; A55182.
 CC DR HSSP; P98066; 1TSG.
 CC DR MGI; MGI:99602; Agcl.
 CC DR InterPro; IPR002353; AntifreezeII.
 CC DR InterPro; IPR007110; Ig-like.
 CC DR InterPro; IPR003006; Ig.MHC.
 CC DR InterPro; IPR003596; Ig.V.
 CC DR InterPro; IPR001304; Lectin_C.
 CC DR InterPro; IPR000538; Link.
 CC DR InterPro; IPR003324; SGXSG.
 CC DR InterPro; IPR000436; Sushi_SCR_CCP.
 CC DR Pfam; PF00047; Ig; 1.
 CC DR Pfam; PF00059; lectin_c; 1.
 CC DR Pfam; PF02339; SGXSG; 60.
 CC DR Pfam; PF00084; sushi; 1.
 CC DR Pfam; PF00193; Xlink; 4.
 CC DR PRINTS; PR01285; LINKMODULE.
 CC DR PRINTS; PR00356; ANTIFREEZEII.
 CC DR ProDom; PD000918; Link; 4.
 CC DR SMART; SM00032; CCP; 1.
 CC DR SMART; SM00034; CLECT; 1.
 CC DR SMART; SM00406; IGV; 1.
 CC DR SMART; SM00445; LINK; 4.
 CC DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 CC DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
 CC DR PROSITE; PS00835; IG LIKE; 1.
 CC DR PROSITE; PS00290; IG.MHC; 1.
 CC DR PROSITE; PS01241; LINK; 4.
 CC KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2132 AGGREGAN CORE PROTEIN.
 FT DOMAIN 34 147 IG-LIKE V-TYPE.
 FT LINK 1.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.
 FT DOMAIN 504 581 LINK 3.
 FT DOMAIN 602 683 LINK 4.
 FT DOMAIN 1918 2044 C-TYPE LECTIN.
 FT DOMAIN 2048 2106 SUSHI.
 FT DOMAIN 48 140 G1-A.
 FT DOMAIN 152 247 G1-B.
 FT DOMAIN 253 349 G1-B'.
 FT DOMAIN 486 580 G2-B.
 FT DOMAIN 587 682 G2-B'.
 FT DOMAIN 685 803 KS.
 FT DOMAIN 805 1231 CS-1.
 FT DOMAIN 1232 1917 CS-2.
 FT DOMAIN 1917 2132 G3.
 FT DISULFID 51 133 BY SIMILARITY.
 FT DISULFID 175 246 BY SIMILARITY.
 FT DISULFID 199 220 BY SIMILARITY.
 FT DISULFID 273 348 BY SIMILARITY.
 FT DISULFID 297 318 BY SIMILARITY.
 FT DISULFID 509 580 BY SIMILARITY.
 FT DISULFID 533 554 BY SIMILARITY.
 FT DISULFID 607 682 BY SIMILARITY.
 FT DISULFID 631 652 BY SIMILARITY.
 FT DISULFID 1922 1933 BY SIMILARITY.
 FT DISULFID 1950 2042 BY SIMILARITY.
 FT DISULFID 2018 2034 BY SIMILARITY.
 FT DISULFID 2049 2092 BY SIMILARITY.
 FT DISULFID 2078 2105 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 26.5222 Seconds
(without alignments)
5244.295 Million cells updates/sec

Title: US-09-688-672A-64
Perfect score: 2737
Sequence: 1 MHHHHTAASDFQLSQGG.....RAPVEADAGGQKVLVRNV 539

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1963	71.7	392	16 Q06267	Q06267 mycobacteri
2	1263	46.1	394	16 Q49722	Q49722 mycobacteri
3	876	24.7	355	16 Q07175	Q07175 mycobacteri
4	474.5	17.3	361	2 Q03020	Q03020 mycobacteri
5	460.5	16.8	354	16 Q9CCV9	Q9CCV9 mycobacteri
6	385.5	14.1	402	16 P96213	P96213 mycobacteri
7	200.5	7.3	446	16 Q8VKA4	Q8VKA4 mycobacteri
8	199.5	7.3	464	16 Q53896	Q53896 mycobacteri
9	188.5	6.9	382	16 Q9CD67	Q9CD67 mycobacteri
10	188.5	6.9	452	2 Q825G6	Q825G6 mycobacteri
11	184.5	6.7	459	16 Q8XQX4	Q8XQX4 raiistonia s
12	180.5	6.6	1468	5 Q9GUB5	Q9GUB5 gallieria me
13	169	6.2	496	16 Q53631	Q53631 mycobacteri
14	169	6.2	533	16 Q8VKR5	Q8VKR5 mycobacteri
15	168	6.1	1417	16 Q9HVN6	Q9HVN6 pseudomonas
16	167	6.1	916	9 Q80116	Q80116 bacterioph

ALIGNMENTS

RESULT 1

Q06267
ID O06267 PRELIMINARY; PRT; 392 AA.
AC O06267;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein KV3616C.
GN RV3616C OR MT3718 OR MTCY07H7B.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

17 166.5 6.1 1018 16 Q9HWU6 pseudomonas
18 163 6.0 536 16 Q8FND6 corynebacte
19 163 6.0 911 16 Q8DXK6 streptococ
20 162.5 5.9 1460 16 Q8K6B1 streptococ
21 161.5 5.9 542 16 Q9FBK9 streptomyc
22 159.5 5.8 810 11 Q8ES29 mus musculu
23 159.5 5.8 860 11 Q8C918 mus musculu
24 159 5.8 1467 16 Q8RY75 deinoococcus
25 159 5.8 3961 16 Q8P942 xanthomonas
26 158 5.8 667 16 Q8XRN8 xanthomonas
27 157.5 5.8 683 16 Q8G791 bifidobacte
28 154 5.6 519 16 Q93J30 streptomyc
29 154 5.6 1329 16 Q06810 mycobacteri
30 153.5 5.6 1488 16 Q8NPX0 corynebacte
31 153 5.6 2751 16 Q8XUK0 raiistonia s
32 152.5 5.6 757 4 Q14234 homo sapien
33 152.5 5.6 1610 16 Q92KQ8 rhizobium m
34 152 5.6 409 16 P95182 mycobacteri
35 152 5.6 628 16 Q8VJ19 mycobacteri
36 151.5 5.5 375 16 Q8DGB7 synchococ
37 151.5 5.5 940 16 Q8K755 bacillus ha
38 151.5 5.5 3145 16 Q98MG7 rhizobium l
39 151.5 5.5 3659 16 Q98LNG rhizobium l
40 151 5.5 525 16 Q8PMV4 xanthomonas
41 151 5.5 588 16 Q50396 mycobacteri
42 151 5.5 1408 16 Q8VK17 mycobacteri
43 150.5 5.5 937 9 Q9G097 lactococcus
44 150.5 5.5 2183 16 Q8PKH6 xanthomonas
45 150 5.5 436 16 Q8VIX6 mycobacteri

DR ENBL; Z95557; CAB08950.1; -;
DR ENBL; A0007171; AAK48077.1; -;
DR TIGR; MT3718; -;
DR TubercuList; RV3616c; -;
KW Hypothetical protein; Complete proteome.
FT CONFLICT 192 192
SQ SEQUENCE 392 AA; 39888 MW; 82BBA8DD9D6F567 CRC64;

Query Match 71.7%; Score 1963; DB 16; Length 392;
Best Local Similarity 99.7%; Pred. No. 2.3e-102;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTISAIDGLYDLIGIGINQGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
Db 1 MSRAFIIDPTISAIDGLYDLIGIGINQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60

QY 208 DKYAGKRNHNVNFFQELADRLQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNVNFFQELADRLQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 120

QY 268 YIPVVGHALSAAPFQACAGAAVVGALAYLVVKTINATOLLKLLAKLAELVAAAIAD 327
Db 121 YIPVVGHALSAAPFQACAGAAVVGALAYLVVKTINATOLLKLLAKLAELVAAAIAD 180

QY 328 IISDVADIIGILGEVWEFITNALNGELWDLKLTGWVTGLFSRGMNSLESFFAGVPGLT 387
Db 181 IISDVADIIGILGEVWEFITNALNGELWDLKLTGWVTGLFSRGMNSLESFFAGVPGLT 240

QY 388 GATSGLSQVTLGFGAAGLSAGSLAHADSLASASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 241 GATSGLSQVTLGFGAAGLSAGSLAHADSLASASLPALAGIGGSGFGGLPSLAQVHAA 300

QY 448 STRQALRPRADGPVGAAGAAVQGSQSVSAQSGQGVGPGVGMGHPSSGASKGTTTKY 507
Db 301 STRQALRPRADGPVGAAGAAVQGSQSVSAQSGQGVGPGVGMGHPSSGASKGTTTKY 360

QY 508 SEGAAAGTDEARAPVEADAGGQKVLVRNVV 539
Db 361 SEGAAAGTDEARAPVEADAGGQKVLVRNVV 392

RESULT 2

Q49722 PRELIMINARY; PRT; 394 AA.
AC Q49722;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein ML0405.
GN ML0405 OR B1620_C2_213 OR MCL383.01.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]_TaxID=1769;
RP SEQUENCE FROM N.A.
RC STRAIN=IN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrall B.G.;
RT "Massive Gene decay in the leprosy bacillus";
RL Nature 409:1007-1011(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

RN [3]
RP SEQUENCE OF 231-394 FROM N.A.
RA Oliver K., Harris D., Parkhill J., Barrall B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 231-394 FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Eigmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
of Mycobacterium leprae";
RL Mol. Microbiol. 7:197-206(1993).
DR EMBL; AL583918; CAC29913.1; -;
DR EMBL; U00015; RAC43223.1; -;
DR EMBL; Z97179; CAB03940.1; -;
DR Leproma; ML0405; -;
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 279 299 POTENTIAL.
SQ SEQUENCE 394 AA; 40754 MW; D0E455C74ED7A10C CRC64;

Query Match 46.1%; Score 1263; DB 16; Length 394;
Best Local Similarity 52.3%; Pred. No. 4e-63;
Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;

QY 148 MSRAFIIDPTISAIDGLYDLIGIGINQGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
Db 1 MSRAFIIDPTISAIDGLYDLIGIGINQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60

QY 208 DKYAGKRNHNVNFFQELADRLQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNVNFFQELADRLQLISLIHQANAVQTTDRDILEGAKKGLFVRPVAVDLT 120

QY 268 YIPVVGHALSAAPFQACAGAAVVGALAYLVVKTINATOLLKLLAKLAELVAAAIAD 327
Db 121 YIPVVGHALSAAPFQACAGAAVVGALAYLVVKTINATOLLKLLAKLAELVAAAIAD 180

QY 328 IISDVADIIGILGEVWEFITNALNGELWDLKLTGWVTGLFSRGMNSLESFFAGVPGLT 387
Db 181 IISDVADIIGILGEVWEFITNALNGELWDLKLTGWVTGLFSRGMNSLESFFAGVPGLT 240

QY 388 GATSGLSQVTLGFGAAGLSAGSLAHADSLASASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 241 GATSGLSQVTLGFGAAGLSAGSLAHADSLASASLPALAGIGGSGFGGLPSLAQVHAA 300

QY 448 STRQALRPRADGPVGAAGAAVQGSQSVSAQSGQGVGPGVGMGHPSSGASKGTTTK 505
Db 301 STRQALRPRADGPVGAAGAAVQGSQSVSAQSGQGVGPGVGMGHPSSGASKGTTTK 360

QY 506 KYSEGAAGTDEARAPVEADAGGQKVLVRNVV 539
Db 361 KYSEGAAGTDEARAPVEADAGGQKVLVRNVV 394

RESULT 3

Q07175 PRELIMINARY; PRT; 355 AA.
ID Q07175
AC Q07175;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 34.9 kDa protein (Serine protease, putative).
GN p5EA OR RV0125 OR MTCI4188.07 OR MT0133.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.",
 RL Nature 393:537-544 (1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.",
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: Z96071; CAB09453.1; -;
 DR EMBL: AE006925; AAK44357.1; -;
 DR TIGR: MT0133; -;
 DR TubercuList; RV0125; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR InterPro; IPR000126; Ser_proteas_V8.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR PRINTS; PRO0839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Hypothetical protein; Serine protease; Protease;
 KW Complete proteome.
 SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 24.7%; Score 676; DB 16; Length 355;
 Best Local Similarity 99.2%; Pred. No. 2.8e-30;
 Matches 132; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HTAASDNFQLSQGGGFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGA 66
 Db :|||||
 223 NTAASDNFQLSQGGGFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGA 282

Qy 67 RVQVWGSAPAAAGISGTGDTAVDGPINSATAMADALNGHPGDVIVTWQTKSGGT 126
 Db :|||||
 283 RVQVWGSAPAAAGISGTGDTAVDGPINSATAMADALNGHPGDVIVTWQTKSGGT 342

Qy 127 RTGNVTLAEGPPA 139
 Db :|||||
 343 RTGNVTLAEGPPA 355

RESULT 4
 Q50320 ID Q50320 PRELIMINARY; PRT; 361 AA.
 AC Q50320;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 34kDa protein precursor.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JD88/107;
 EX MEDLINE=95005449; PubMed=7921248;
 RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
 RT "Identification and characterisation of a putative serine protease

RT expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
 RL Microbiology 140:1977-1982 (1994).
 CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 DR EMBL: Z23092; CA80638.1; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Signal.
 FT SIGNAL 1 38 POTENTIAL.
 SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match 17.3%; Score 474.5; DB 2; Length 361;
 Best Local Similarity 69.7%; Pred. No. 5.6e-19;
 Matches 92; Conservative 17; Mismatches 22; Indels 1; Gaps 1;

Qy 8 TAASDNFQLSQGGGFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGAR 67
 Db :|||||
 231 TAATDSYKMS-GGQGFAPIGMAVAVANQIRSGAGSNVHIGPTAFGLGVVDNNGAR 289

Qy 68 VQVWGSAPAAAGISGTGDTAVDGPINSATAMADALNGHPGDVIVTWQTKSGGTR 127
 Db :|||||
 290 VQVWVTGPAAGIAPGDVITGVTVPINGATSNTEVLVPHHPGDITIAVHRSVDGGER 349

Qy 128 TGNVTLAEGPPA 139
 Db :|||||
 350 TANITLAEQPPA 361

RESULT 5
 Q9CCY9 ID Q9CCY9 PRELIMINARY; PRT; 354 AA.
 AC Q9CCY9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Probable secreted serine protease.
 GN ML2659.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.",
 RL Nature 409:1007-1011 (2001)
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AL589926; CAC32191.1; -;
 DR Leproma; ML2659; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR InterPro; IPR000126; Ser_proteas_V8.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR PRINTS; PRO0839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 1.

```

DR PROSITE; PS50106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9EA4A CRC64;

Query Match 16.8%; Score 460.5; DB 16; Length 354;
Best Local Similarity 66.2%; Pred. No. 3.3e-18;
Matches 88; Conservative 18; Mismatches 26; Indels 1; Gaps 1;

QY 7 HTAASNFOQLSQGGQFAIPGQMAIAQIRSGGSPVHIGPTAFGLGVVDNNGNGA 66
DQ 223 NTAATDNYKM-LGGQFAIPGQMEWGAIRSGAGSNTVHIGPTAFGLGLVDNNGNGA 281
QY 67 RVORVVGSAAPAGSLGSDGVTAVGAPINSGATAMADALNGHPCDVISVTWQKSGT 126
DQ 282 RVARVATGPAAGVAGISVGDITSVDGVPISATANTNLVPHHPGCTVAVNRSAGGSD 341
QY 127 RTGNVTLAEGPPA 139
DQ 342 LTANVTLAEGPPA 354

RESULT 6
P96213 ID P96213 PRELIMINARY; PRT; 402 AA.
AC P96213;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Hypothetical protein Kv3864.
GN RV3864 OR MTCY01A6.04C OR MT3978.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z83864; CAB06237.1; -.
DR EMBL; A5007189; AAK48347.1; -.
DR TIGR; MT3978; -.
DR TubercuList; Kv3864; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 402 AA; 42068 MW; B400E0E22D482765 CRC64;

Query Match 14.1%; Score 385.5; DB 16; Length 402;
Best Local Similarity 30.4%; Pred. No. 6.3e-14;
Matches 130; Conservative 47; Mismatches 156; Indels 95; Gaps 13;

QY 157 TISAIDGLYDLGIPNGGGILYSLEYFEKALEBELAAAFPGDWLGSAAADKYAGKNRN 216

```

```

Db 11 TSNFIWQLLLLLGGEGIDPDGDIINTGSSLFQISDRKGLAIPGTNNWIGQAAEAYLNQNI 70
QY 217 HVNFFQELADLDROLISLIHDQANAVQTTDRDILEGAKGLEGFVRPVAVDLTYIPVVGCHAL 276
DQ 71 QQLRAQVWGDLKLTGNWISNQAKYVSDTEFVLRAMKMDGVYKVKGLEKIFLLGLHW 130
QY 277 SAFAQAFPCAGANAVGAGLAYLVVTKLINATOLLKLLAKLAEVAAAIADIISDVADII 336
DQ 131 SWELAIEMSGIAVAVVGALLYLTIMTNATNL- 164
QY 337 KGIIGVWEIITN-----ALNGKELWDLKLTG-W-----VTGLFSRCW--- 373
DQ 165 RGLIGRLIEMLTLPKPPGLPGLPSLDIDGLWPKLPDIPPGLPDIPGLDFKPPPT 224
QY 374 -----SNLSPF-----FAGVPLGTATSGLSQVTLG-F-GAAGL- 405
DQ 225 FGSPLFPDLPSPFGFPFEPFAIPGPP-ALPGLPSIPNLPGLPGLGLDLLPGVGLGKL 283
QY 406 -SASSGLAHADSLASSASLPALAGIGGSGFGLPSLAQVHAASLQALRPADGPVGA 464
DQ 284 PTWELAAALPDFLGGFAGLPSL-FGNNLSFASLPTVQGTATMGQQLQVAAAGGSPSQL 342
QY 465 AEQVGGSQLVSAQSGQGMGPVGMGMHPSSGASKGTTTKYSEGAAGTDAERAPVE 524
DQ 343 ASMSGQQAQLISSQAQGG-----GQGHATLVSDK-----KEDEGVA---EERAPID 387
QY 525 ADAGGGQK 532
DQ 388 AGTAASQR 395

RESULT 7
Q8VKA4 ID Q8VKA4 PRELIMINARY; PRT; 446 AA.
AC Q8VKA4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Heat shock protein HtrA, putative.
GN M1011.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006985; AAK45259.1; -.
DR TIGR; MT1011; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00535; PDZ_1.
DR Pfam; PF00089; tryptophan; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
SQ SEQUENCE 446 AA; 44484 MW; 54170CBEA8F872B CRC64;

Query Match 7.3%; Score 200.5; DB 16; Length 446;
Best Local Similarity 38.9%; Pred. No. 0.0017;
Matches 51; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 8 TAASDNFQLSQG--GGFAIPGQMAIAQIRSGGSPVHIGPTAFGLGVVDNNGN 64

```

```
Db 319 TLGADSADAQSGSIGLGLFAIPVDQAKRIADELISTGKA-----SHASLGVQVINDKDT 372
QY 65 GARVQRVVGSAAPASLSIGSTGVDITAVDGAIPINSATAMADALNGHHPCDVISVTWQTKSG 124
D 373 GAKIVEVVGAAANAGVPKGVVTVKVDPRINSADALVAARSKAPGATVALTFQDPSG 432
QY 125 GTRTGNVTLAE 135
D 433 GSRTVQVTLGK 443

RESULT 8
O53896 PRELIMINARY; PRT; 464 AA.
AC O53896;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative serine protease.
GN RV0983 OR MTW044.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL021999; CAA17582.1; -.
DR TubercuList; RV0983; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 464 AA; 46452 MW; AE93BFCCS3E1EC8F CRC64;

Query Match 7.3%; Score 199.5; DB 16; Length 464;
Best Local Similarity 38.9%; Pred. No. 0.002;
Matches 51; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 8 TAASNFQLSQ--CGGAIPIGQAMAIAGQIRSGGSPVTHIGTAPLGLGVV-DNNGN 64
D 337 TLGADSADAQSGSIGLGLFAIPVDQAKRIADELISTGKA-----SHASLGVQVINDKDTL 390
QY 65 GARVQRVVGSAAPASLSIGSTGVDITAVDGAIPINSATAMADALNGHHPCDVISVTWQTKSG 124
D 391 GAKIVEVVGAAANAGVPKGVVTVKVDPRINSADALVAARSKAPGATVALTFQDPSG 450

QY 125 GTRTGNVTLAE 135
D 451 GSRTVQVTLGK 461

RESULT 9
Q9CD67 PRELIMINARY; PRT; 382 AA.
ID Q9CD67
```

```
Q9CD67;
AC 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Possible secreted serine protease.
GN ML0176.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrall B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL58917; CAC29684.1; -.
DR Leproma; ML0176; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 382 AA; 37084 MW; 3DD8DD8AE32A80D CRC64;

Query Match 6.9%; Score 188.5; DB 16; Length 382;
Best Local Similarity 39.8%; Pred. No. 0.0065;
Matches 51; Conservative 19; Mismatches 51; Indels 7; Gaps 3;

QY 9 AASDNFQLSQGGQGAIPIGQAMAIAGQIRSGGSPVTHIGTAPLGLGVV-DNNGN-GAR 67
D 258 ADSGDAQSGSIGLGLFAIPVDQAKRIADELISTG--KATH----ASLGVQVATDKGTPGAK 311
QY 68 VQRVVGSAAPASLSIGSTGVDITAVDGAIPINSATAMADALNGHHPCDVISVTWQTKSGGTR 127
D 312 VMDVAGGAANAANAAPKGVVTVKVDRLISSADALVAARSKAPGDKVSLTYQDQSGSR 371

QY 128 TGNVTLAE 135
D 372 TVQVTLGK 379

RESULT 10
Q9Z5G6 PRELIMINARY; PRT; 452 AA.
ID Q9Z5G6
AC Q9Z5G6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative serine protease.
GN MLCB373.28.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris D., Taylor K.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
```

SEQUENCE FROM N.A.
 RA James K.D., Parkhill J., Barrell B.G., Rastam M.A.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN [3].
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93188700; PubMed=8446027;
 RA Englemer K., Honore N., Woods S.A., Caudron B., Cole S.T.;
 RT "Use of an ordered cosmid library to deduce the genomic organization
 of *Mycobacterium leprae*.";
 RL Mol. Microbiol. 7:197-206(1993).
 DR EMBL; AL035500; CAB36690.1; --
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR Pfam; PF00595; PDZ; 1.
 DR PRINTS; PR00089; tryptsin; 1.
 DR SMART; SMC0228; PDZ; 1.
 KW Hydrolase; Protease, Serine protease.
 SQ SEQUENCE 452 AA; 45130 MW; 6CA675BE0911F983 CRC64;
 Query Match 6.9%; Score 188.5; DB 2; Length 452;
 Best Local Similarity 39.8%; Pred. No. 0.0081;
 Matches 51; Conservative 19; Mismatches 51; Indels 7; Gaps 3;
 QY 9 AASDNFQLSCGGGFAIPICQAMAIAGQIRSGGSPVHIGTAFGLGVVDNNGN-GAR 67
 DB 328 ADSGDAQSGSIGLFAIPVDQAKRADELISG--KATH---ASLGVQVATDKTGPAG 381
 QY 68 VQWVGSAAPASIGSTGVITAVDCAPINSATAMADALNGHHPGDVISTVWTKSGGTR 127
 DB 382 VMDVAGGAANAAPKGVVLTKYVDRDLISSADALVAARSKAPGKXSLTQDQSGSSR 441
 QY 128 TGNVTIAE 135
 DB 442 TVQVTLGK 449
 RESULT 11
 Q8XQX4 PRELIMINARY; PRT; 459 AA.
 ID Q8XQX4
 AC Q8XQX4
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Putative hemagglutinin-related transmembrane protein.
 GN RSP1094 OR RS02600.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Ralstoniaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*.";
 RL Nature 415:497-502(2002).
 DR EMBL; AL646082; CAD18245.1; --
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 459 AA; 41117 MW; FE06C4144483ACC9 CRC64;
 Query Match 6.7%; Score 184.5; DB 16; Length 459;
 Best Local Similarity 24.7%; Pred. No. 0.014;
 Matches 133; Conservative 60; Mismatches 200; Indels 145; Gaps 28;
 QY 32 AIAGQIRSGG---GSPTVHIGTAFGLGVVDNNGNARVQVVGSAAPASLIGSTGVD 87

DB 13 AVAGLILIGGCASSGS-----GDTSGT-LGSSSSNGA--SGSAGGGSSGSDTSGTSGTS 64
 QY 88 ITAVDGAPINSATAMADALNGHHPGDVISTVWTKSG-GTRTGNVTIAEGPPAEFLVPRG 146
 DB 65 GTNGSGTTPLTAA-GTVLN--NSGGVVTTAGTSLGIGTSLKNANLP-----LVPSS 114
 QY 147 SMSRAFIIDPTISAIIDLGLIGIPNQGILYSSLEVEFEKALEELAAAPPGDGLGSA 206
 DB 115 AQ-----SGLGGVVENLGATVSALG-----AGVDG-LGS- 143
 QY 207 ADYAGKRNHNVFFQELADRLQLISLIHQANAVOTTRDIIEGAKKG-LEFVRPVAVD 265
 DB 144 ---IGSNPNPIG--TTVASTG---NVVTGVNTVTSAGTLVGLGTGQLSLAPVTP 193
 QY 266 LTYIPVVGHALSAAFQAFPCAGAMVVGAL--AYLVVKTLINATQLLKLAKLAELVAA 323
 DB 194 L-----AGALTQVQAVTVAGTTLGTTLSTGTPVEQLTQQLSTAIVP 234
 QY 324 AIADIISDVADIKILGEVMEFITNALNGLKELWDLKLTGVTGTLFSGRWSNLESFFAGV 383
 DB 235 LTSQIIGTQTV-----GATTCWGTTPANSLTGLGAIANGSTLTN--ANV 279
 QY 384 PGLTGATSGLSQVTLGFCAGLSASS--GLAHADSLASSASL-PALAGIGGSGGFGGLPS 440
 DB 280 P-----VVSNGVGVVAVGNTVASLGGVAYSASSASPLAPVTCALGGGNPLGAVTS 332
 QY 441 LAQVHAASTRQALRPADGPFVGAALAEQVG-----QSOLVSAQSGQMGQGPV-- 487
 DB 333 VLGLGSGTSG-SPLA--FVTGVLSSVGATGGSSPLAAVTSTLGLGAAAGSGSPLAP 399
 QY 488 ---GMGMHRESSGASKGTTTKYSEGAACTEDA-----ERAPVEADAGG 529
 DB 390 VTSALGSGVGTGCG--GTLAPVASITAPVGTVATVGTSTLTGATTPAPVTGSAGG 445
 RESULT 12
 Q9GUB5 PRELIMINARY; PRT; 1468 AA.
 ID Q9GUB5
 AC Q9GUB5
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Heavy-chain fibroin (fragment).
 GN FIB-H.
 OS Galleria mellonella (Wax moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Pyraloidea;
 OC Pyralidae; Galleriinae; Galleria.
 OX NCBI_TaxID=7137;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Posterior silk gland;
 RA Zurovec M., Kodrik D., Yang C., Sehna F.;
 RT "Heavy-chain fibroin of *Galleria mellonella* L.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF095239; AAC10393.1; --
 FT NON TER 1468 1468
 SQ SEQUENCE 1468 AA; 122705 MW; 2DA59E1181B3DDF CRC64;
 Query Match 6.6%; Score 180.5; DB 5; Length 1468;
 Best Local Similarity 25.1%; Pred. No. 0.11;
 Matches 138; Conservative 42; Mismatches 219; Indels 151; Gaps 22;
 QY 28 GQAMATAGQIRSGGSPVHIGTAFGLGVVDNNGNARVQVVGSAAPASLIGSTGVD 87
 DB 473 GVSTAGSLGGUGGAGVAVGP-AGAGLGGVAGGSGS-----SAASAARSAPVP 526
 QY 88 ITAVDGAPINSATAMADA-----LNHHHPGDVISTVWTKSGGTRTG---- 129
 DB 527 IVIEDGSSAASAAAAGSGASGLGLGAWGPLGGIGPNEVSSAS-ATGSAAGSTGAGLG 585
 QY 130 -----NVTLAGPPAEFLVPRGMSRAFIIDPTISAIIDLGLIGI---PNQ 175

```
Db 586 GSGAAGSAAAGCAAGPAPVIVIEDGSSAAS-----AAAAGSAGSLGGLGAGWGT 640
Qy 176 GGILYSLEYEFALEELAAAFPGDGLGSAADRYAGKNRHNHNFQELADLPQLISLI 235
Db 641 GGIGPNQVS--SASATGSAAGSIGSLGCGSA--AGSS-----674
Qy 236 HDQANAVTTTRDILEGAKGLLEFVRPVAVDLTYIPVGHASLAAPFAFPAGANAVVG-- 293
Db 675 -----LASAASGAAGAAPVI-----VIEDGSSAASAAAAGSAGSGVGLG 714
Qy 294 -GALAYLVVKTINATQLLKLAELVAAAADIISDVADI--KGIIGWEWETITNALN 352
Db 715 LGALGPL-----GGIGPIGASSAGSAGAGLGGVGAAGTSGLG 751
Qy 353 GKELWMDKLTGWTLFRSGWSNLESFPAGVPL-TGATGSLGSLVTLGFAAG-----L 405
Db 752 GI-----GCV-----GASTAGSAGAGLGGIGAGGSSGSSAASAASGASGAGEVIVI 797
Qy 406 SASSGLAHADSLASSASIPALAGIGGGSGFGLPFLAQVHAASRQALRPADGPGVAAA 465
Db 798 DRSAAASAAAAGSAGSGFGLGVLGWPGLGIGPIGASSASAGL-----GCVGAAG 852
Qy 466 EQ-----VGGQSOLVAQSQMG-CPVGMGMHPFSGAKGTTTTKXSEGAAGTDAER 520
Db 853 TSGLGGLGAG--ASAAGSAGAGLGGIGAGGSSGSSVASAASGTS--GAGEVIVIDRSS 908
Qy 521 APVEADAGG 530
Db 909 AASAAAAGSG 918

RESULT 13
OS3631 PRELIMINARY; PRT; 496 AA.
ID OS3631
AC OS3631
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE PGSR-family protein.
GN RV0109 OR MT031.03C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Kroch A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL Nature 393:537-544(1998).
DR EMBL; AL021926; CAA17303.1; -.
DR Tuberculist; rv0109; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Complete proteome.
SQ SEQUENCE 496 AA; 42289 MW; A3FLAD215EC11786 CRC64;

Query Match 6.2%; Score 169; DB 16; Length 496;
Best Local Similarity 25.5%; Pred. No. 0.11;
Matches 83; Conservative 26; Mismatches 107; Indels 110; Gaps 14;

Qy 262 VAVDLTYIPVGHALS---AFAQAPFCAGAMAVG-----293
```

```
Db 11 VAAATHLAGISALSTANAAAAP--TTALSVAGADEVSVLIAALFEAYAEYQALSQA 68
Qy 294 -----GALAYLVVKTINATQLLKLAELVAAAADIISVDADI---336
Db 69 ALAFHQDFVOALNMGAVCYAAAEAT-ANATPL-----QALQTVQONVLTIVNAP 115
Qy 337 -KGIIGWEWETITNALNGKLWDLKLTGWVTG-----LFSRGSNLSFPAGVPLGTGATS 391
Db 116 TQALLGR--PIITNGANGLPN-----TGODGPGGLLFNGGN-----GSG 155
Qy 392 GLSQVTGLFGAAGL--SASSGLAHADSLASSASLPA-----LAGIGGSGFGLPFLAQV 444
Db 156 GVDQAGNGGAAGLIGNGGSGVGGPGIAGSAGGAGAGLFFNGGPGGAGGIGITGD- 214
Qy 445 HAASTQALLRPADGVPVGAAPVQVQSQSOLVAQSQMGVPGVGMGHPSSGASKGTTT 504
Db 215 -----GGPGAGGNAIG---LFGSGGTGGMGGVGGMGVGNAGNGGTA 257
Qy 505 KYSEGAAGTDAERAPVEADAGG 530
Db 258 GLFGHGAGAGGAGISADGGLGGGG 283

RESULT 14
Q8VKR5 PRELIMINARY; PRT; 533 AA.
ID Q8VKR5
AC Q8VKR5
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE PE PGSR family protein.
GN MT0118.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006923; AAK4341.1; -.
DR TIGR; MT0118; -.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSHLL.
DR ProDom; PD001223; PE_region; 1.
SQ SEQUENCE 533 AA; 46785 MW; 56F0115F01F4D4E5 CRC64;

Query Match 6.2%; Score 169; DB 16; Length 533;
Best Local Similarity 26.2%; Pred. No. 0.13;
Matches 89; Conservative 25; Mismatches 134; Indels 92; Gaps 15;

Qy 229 RQLISLHDOANAVOTTRDILEGAKKLEFVR-PVAVDLTYIPVVG-----273
Db 35 RSLMSLITSPATVAAAATHLAGISALSTANAAAAPTALSVAGADEVSVLIAALFEA 94
Qy 274 -----HALSA---AFAQAPFCAGAMAVVGGALAYLVVKTINATQLLKLAELVAAA 325
Db 95 YAEYQALSQAALAFHQDF---VQALNMGAVCYAAAEAT-ANATPL-----QAL 138
Qy 326 ADIISVDADI-----KGIIGWEWETITNALNGKLWDLKLTGWVTG-----LFSRGSNLS 377
Db 139 QTVQONVLTIVNAPTQALLGR--PIITNGANGLPN-----TGODGPGGLLFNGGN--- 188
```

```
QY 378 SFAGVPLTGATSGLSQVLTGLEAGL--SASSGLAHADSLASSASLPA-----LAGIG 430
Db 189 -----GGGGVDQAGNGGAAAGLNGGGGGVGGPGIAGSAGGAGGLLFNG 238
QY 431 GSGGFGPLSLAQVHAASRQALRPRADGVRGAAAEQVGGQSOLVSAQSGQGGGPGVNG 490
Db 239 GPGAGGIGTGD-----GGPGAGGNAIG-----LFGSGGTGNGGPGVNG 280
QY 491 GMPSSGASGTTTKYSEGAAGTDAERAPVEADAGG 530
Db 281 GVGNGGNAGNGGTAGLFGHGAGGAGGIGSADGGLGGGG 320

RESULT 15
Q9HVN6 PRELIMINARY; PRT; 1417 AA.
AC Q9HVN6;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein PA4541.
GN PA4541.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=2043737; PubMed=10994043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004867; AGO7929.1; ..
DR HSSP; P22629; 1SWC.
DR InterPro; IPR006162; Ppantne attach.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1417 AA; 139958 MW; 212C916D5A55C39D CRC64;
```

```
Query Match 6.1%; Score 168; DB 16; Length 1417;
Best Local Similarity 22.1%; Pred.No.0.55;
Matches 139; Conservative 77; Mismatches 206; Indels 206; Gaps 31;

QY 19 GGGFAIPIQQAVAIAG-----QIRSGG 41
Db 782 GGNSSFLSICNASAFGGTDFGLGNTIDNLAVYGTGAYSLFVNRGTLNMLNLERISADG 841
QY 42 GSPT---VHIGPTAFGLGVNDN-NGNGARVQVVGSPARSIGISTGDIVTAVDGPIN 97
Db 842 AQTHNVQVGSUAAVNLGRIDNVNASDIRI-----AAASKNSLGG--LVALNLGSD 893
QY 98 SATAMADALNGHHPGDVI SVTWQTSKGGTRTGNVTIAEG-----PPAEFLVP----- 144
Db 894 NASASGTLVGNRH-----TYAL--GGLAENISTARCVASISNSRADFAISGOLKQDA 944
QY 145 -----RGSMSRAFIIDPTISAIDGLYDLIGIPIQGGIL-YSSLEYFEKALE 191
Db 945 SHYGAGGLVNRNGGLIRS-----SGSQGTLSLSGHGM-NLGLGVYSS-----AGGLA 992
QY 192 ELAAA--FFGDGWLGSADKYAGKNRNVNFFQELADLDROLISLIHQANAV--QTRD 247
Db 993 DVASVDVSGNGQRG---LYGGLICLVN-----SGIAHATASKVGTDAE 1036
QY 248 ILEGAKKLEFVRPNAVDITYPVVCHALSAAFAFPFAGAMAV-----VCGALAY----- 298
Db 1037 ALGG-----LIGNLNAAINNAAHGDVSLQAGRYLGLGIGHNQAG 1077
```

```
QY 299 -----LVVKTLLINATQLLKLAKLAEVLVAAATADIISDVADIKIGIL----- 340
Db 1078 NLNVSTSGNLGGSLIQAGGLIGLNANASLVNASAKGNVATRGAEAVGGLIGENLYGSV 1137
QY 341 -----GEVNEFITNALNGLKELWDKLTGVVTCGLFSEGMNLSFFAGVPLGTGATSGLS 394
Db 1138 INGSASGEVTDGSGKTIGGL--IGSNLGGNHSNLIKASGWN-----AGAN---S 1181
QY 395 QVTGLFQAAGLSASSGLAHADSLASSASLPALAG--ICGGSGFGGLPGLSLAQVHAASRQA 452
Db 1182 DVGGLIG-----HNRGNGHS-TLAASGNVTGGKSRVGLGVNDAAASLTVNASGNVSA 1235
QY 453 LRPRADGVRGAAAEQVGGQSOLVSAQ-----SQNGGPGVNG-----GMHSSGASKG 501
Db 1236 SGSEAIG--GLIGSDLRGSLMLASSHGIVNDKTSNHLGGLVGRGENTSIRSASKASGAVSG 1293
QY 502 TTTKKYSEGAAGTDAERAPVEADAGG 529
Db 1294 GAGIR-AGGLVGSLECHQALILGASAGG 1320
```

Search completed: November 21, 2003, 16:08:18
Job time : 27.5222 secs